IBM SPSS Statistics Base 32

Nota

Antes de utilizar esta información y el producto al que da soporte, lea la información del apartado “Avisos” en la página 323.

Información del producto

Esta edición se aplica a la versión 32, release 0, modificación 0 de IBM® SPSS Estadística y a todas las versiones y modificaciones posteriores hasta que se indique lo contrario en nuevas ediciones.

© Copyright International Business Machines Corporation .

Contenido

Capítulo 1. Características principales..... 1
Análisis de potencia..... 1
Medias..... 2
Proporciones..... 16
Correlaciones..... 27
Regresión..... 35
Análisis de potencia: Precisión..... 38
Análisis de potencia: Valores de cuadrícula..... 39
Metaanálisis..... 39
Meta-análisis continuo..... 40
Tamaño de efecto de meta-análisis continuo..... 48
Meta-análisis binario..... 56
Tamaño de efecto de meta-análisis binario..... 65
Metaanálisis Regresión..... 73
Opciones de trazado de meta-análisis..... 78
Libro de códigos..... 84
Pestaña Resultados de libro de códigos..... 85
Pestaña Estadísticos del libro de códigos..... 87
Frecuencias..... 87
Frecuencias: Estadísticos..... 88
Frecuencias: Gráficos..... 89
Frecuencias: Formato..... 90
Descriptivos..... 90
Descriptivos: Opciones..... 91
Características adicionales del comando DESCRIPTIVES..... 91
Percentiles..... 92
Percentiles de valores perdidos..... 92
Explorar..... 92
Explorar: Estadísticos..... 93
Explorar: Gráficos..... 94
Explorar: Opciones..... 95
Características adicionales del comando EXAMINE..... 95
Tablas cruzadas..... 95
Capas de las tablas cruzadas..... 96
Gráficos de barras agrupadas..... 97
Tablas cruzadas mostrando variables de capa en capas de tabla..... 97
Estadísticos de tablas cruzadas..... 97
Visualización en casillas de tablas cruzadas..... 99
Formato de tablas cruzadas..... 100
Resumir..... 100
Resumir: Opciones..... 101
Resumir: Estadísticos..... 101
Medias..... 102
Medias: Opciones..... 103
Cubos OLAP..... 105
Cubos OLAP: Estadísticos..... 105
Cubos OLAP: Diferencias..... 107
Cubos OLAP: Título..... 107
Proporciones..... 107
Introducción a las proporciones..... 107
Proporciones de una muestra..... 108

Proporciones de muestras emparejadas....111
Proporciones de muestras independientes.... 113
Pruebas T.... 116
Pruebas T....116
Prueba T para muestras independientes.... 117
Prueba T de muestras emparejadas....118
Prueba T para una muestra....120
Características adicionales del comando T TEST....121
ANOVA de un factor.... 121
ANOVA de un factor: Contrastes....123
ANOVA de un factor: Contrastes post hoc....123
ANOVA de un factor: Opciones.... 125
Características adicionales del comando ONEWAY....126
GLM Análisis univariante.... 126
GLM:Modelo.... 128
GLM:Contrastes....129
Gráficos de perfil GLM....130
GLM:Comparaciones post hoc.... 132
MLG: Guardar....135
Medias marginales estimadas GLM.... 136
Opciones MLG....136
Características adicionales del comando UNIANOVA.... 138
Bland-Altman....138
Bland Altman: Criterios....139
Bland Altman: Imprimir.... 139
Bland Altman: Argumento....139
Prueba de chi cuadrado....139
Prueba Chi Cuadrado: Exacto.... 140
Prueba Chi Cuadrado: Células.... 140
Prueba Chi Cuadrado: Formato....141
Correlaciones bivariadas.... 141
Correlaciones bivariadas: Opciones.... 142
Intervalo de confianza de correlaciones bivariadas.... 143
Características adicionales de los comandos CORRELATIONS y NONPAR CORR.... 143
Correlación de proximidad.... 143
Cartografía de proximidad: Datos....145
Cartografía de proximidad: Pesos....146
Cartografía de proximidad: Inicio.... 147
Cartografía de proximidad: Modelo.... 147
Cartografía de proximidad: Restricciones.... 149
Cartografía de proximidad: Suplemento.... 150
Cartografía de proximidad: Criterios.... 151
Cartografía de proximidad: Salida.... 152
Cartografía de proximidad: Guardar....158
Correlaciones parciales....158
Correlaciones parciales: Opciones.... 159
Características adicionales del comando PARTIAL CORR.... 160
Distancias....160
Distancias: Medidas de disimilaridad....160
Distancias: Medidas de similaridad.... 161
Características adicionales del comando PROXIMITIES....161
Correlación de distancias.... 161
Correlación de distancia: Criterios....162
Correlación de distancia: Imprimir....163
Correlación de distancias: Gráfico....164
Modelos lineales....164
Para obtener un modelo lineal....165
Objetivos 165

Conceptos básicos ... 165
Selección de modelos ... 166
Conjuntos ... 167
Avanzado ... 167
Opciones de modelos ... 167
Resumen del modelo ... 168
Preparación automática de datos ... 168
Importancia de predictor ... 168
Predicho por observado ... 169
Residuos ... 169
Valores atípicos ... 169
Efectos ... 169
Coeficientes ... 170
Medias estimadas ... 170
Resumen de generación de modelos ... 170
Regresión lineal ... 171
Métodos de selección de variables en el análisis de regresión lineal ... 172
Regresión lineal: Establecer regla ... 173
Regresión lineal: Gráficos ... 173
Regresión lineal: almacenamiento de variables nuevas ... 173
Regresión lineal: Estadísticos ... 175
Regresión lineal: Opciones ... 175
Características adicionales del comando REGRESSION ... 176
Regresión ordinal ... 176
Regresión ordinal: Opciones ... 177
Resultado de la regresión ordinal ... 178
Modelo de ubicación de la regresión ordinal ... 179
Modelo de escala de la regresión ordinal ... 179
Características adicionales del comando PLUM ... 180
Regresión lineal de Elastic Net ... 180
Regresión de red elástica lineal: Opciones ... 181
Regresión lineal de Lasso ... 184
Regresión de lazo lineal: Opciones ... 184
Regresión lineal Ridge ... 187
Regresión de cresta lineal: Opciones ... 188
Análisis de mediación ... 190
Tipo de modelo ... 195
Variables dependientes ... 196
Predictor ... 197
Criterios ... 198
Imprimir ... 200
Gráficos ... 201
Filtros de series temporales ... 202
Filtros de series temporales: Opciones ... 203
Estimación curvilínea ... 204
Modelos del procedimiento Estimación curvilínea ... 205
Estimación curvilínea: Guardar ... 205
Regresión por mínimos cuadrados parciales ... 206
Modelo ... 207
Opciones ... 207
Análisis de vecinos más cercanos ... 208
Vecinos ... 210
Características ... 210
Particiones ... 211
Guardado ... 212
Resultado ... 212
Opciones ... 212
Vista de modelo ... 213

Análisis discriminante ... 216
Análisis discriminante: Definir rango ... 217
Análisis discriminante: Seleccionar casos ... 217
Análisis discriminante: Estadísticos ... 217
Análisis discriminante: Método de inclusión por pasos ... 218
Análisis discriminante: Clasificar ... 218
Análisis discriminante: Guardar ... 219
Características adicionales del comando DISCRIMINANT ... 219
Análisis factorial ... 220
Selección de casos en el análisis factorial ... 221
Análisis factorial: Descriptivos ... 221
Análisis factorial: Extracción ... 221
Análisis factorial: Rotación ... 222
Análisis factorial: Puntuaciones factoriales ... 223
Análisis factorial: Opciones ... 223
Características adicionales del comando FACTOR ... 223
Selección de procedimientos para la agrupación en clústeres ... 224
Análisis de clústeres bietápicos ... 224
Opciones del análisis de clústeres en dos fases ... 226
Resultados de análisis de clústeres en dos fases ... 227
El visor de clústeres ... 227
Análisis de clústeres jerárquico ... 232
Análisis de clústeres jerárquico: Método ... 233
Análisis de clústeres jerárquico: Estadísticos ... 234
Análisis de clústeres jerárquico: Gráficos ... 234
Análisis de clústeres jerárquico: Guardar variables nuevas ... 234
Características adicionales de la sintaxis de comandos CLUSTER ... 234
Análisis de clústeres de K-medias ... 234
Eficacia del análisis de clústeres de K-medias ... 235
Análisis de clústeres de K-medias: Iterar ... 236
Análisis de clústeres de K-medias: Guardar ... 236
Análisis de clústeres de K-medias: Opciones ... 236
Características adicionales del comando QUICK CLUSTER ... 237
Pruebas no paramétricas ... 237
Pruebas no paramétricas de una muestra ... 237
Prueba de Wilcoxon de una muestra ... 241
Pruebas no paramétricas de muestras independientes ... 242
Pruebas no paramétricas de muestras relacionadas ... 244
Características adicionales del comando NPTESTS ... 247
Cuadros de diálogo antiguos ... 247
Análisis de respuestas múltiples ... 258
Análisis de respuestas múltiples ... 258
Definir conjuntos de respuestas múltiples ... 259
Frecuencias de respuestas múltiples ... 260
Tablas cruzadas de respuestas múltiples ... 261
Informes de los resultados ... 262
Informes de los resultados ... 262
Resúmenes de informes en filas ... 263
Resúmenes de informes en columnas ... 265
Características adicionales del comando REPORT ... 267
Análisis de fiabilidad ... 268
Análisis de fiabilidad: Estadísticos ... 269
Características adicionales del comando RELIABILITY ... 271
Kappa ponderada ... 272
Kappa ponderada: Criterios ... 273
Kappa ponderada: Imprimir ... 273
Escalamiento multidimensional ... 274
Escalamiento multidimensional: Forma de los datos ... 275

Escalamiento multidimensional: Crear la medida a partir de los datos....275
Escalamiento multidimensional: Modelo....276
Escalamiento multidimensional: Opciones....276
Características adicionales del comando ALSCAL....276
Comparación entre ALSCAL y PROXMAP....276
Estadísticos de la razón....282
Estadísticos de la razón: Estadísticos....282
Gráficos P-P....284
Gráficos Q-Q....286
Análisis ROC....289
Análisis ROC: Opciones....290
Análisis ROC: Visualización....291
Análisis ROC: Definir grupos (serie)....291
Análisis ROC: Definir grupos (numérico)....292
Curva ROC....292
Curvas ROC: Opciones....293
Simulación....293
Para diseñar una simulación basada en un archivo de modelo....294
Para diseñar una simulación basada en ecuaciones personalizadas....294
Diseñar una simulación sin un modelo predictivo....295
Para ejecutar una simulación de un plan de simulación....295
Generador de simulaciones....296
Cuadro de diálogo Ejecutar simulación....307
Trabajo con resultados de gráficos de simulación....309
Modelado geoespacial....311
Selección de mapas....311
Orígenes de datos....313
Reglas asociación geoespacial....314
Predicción temporal espacial....318
Finalizar....321

Avisos....323
Marcas comerciales....324

Índice....327

Capítulo 1. Características principales

Las características principales siguientes están incluidas en IBM SPSS Estadística edición Base.

Análisis de potencia

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

IBM SPSS Estadística proporciona los siguientes procedimientos de Power Analysis:

Prueba T de una muestra
En el análisis de una muestra, los datos observados se recopilan como una única muestra aleatoria. Se supone que los datos de la muestra siguen de forma independiente e idéntica una distribución normal con una media y una varianza fija y extraen una inferencia estadística sobre el parámetro de la media.

Prueba T de muestra emparejada
En el análisis de muestras emparejadas, los datos observados contienen dos muestras emparejadas y correlacionadas y cada caso tiene dos mediciones. Se supone que los datos de cada muestra siguen de forma independiente e idéntica una distribución normal con una media y una varianza fija, y extraen una inferencia estadística sobre la diferencia de las medias.

Prueba T de muestra independiente
En el análisis de muestras independientes, los datos observados contienen dos muestras independientes. Se supone que los datos de cada muestra siguen de forma independiente e idéntica una distribución normal con una media y una varianza fija, y extraen una inferencia estadística sobre la diferencia de las medias.

ANOVA unidireccional
El análisis de varianza (ANOVA) es un método estadístico para realizar una estimación de las medias de varias poblaciones que, a menudo, se supone que se distribuyen normalmente. El ANOVA de un factor, un tipo común de ANOVA, es una extensión de la prueba *t* de dos muestras.

Ejemplo. La potencia de un contraste de hipótesis para detectar una hipótesis alternativa correcta es la probabilidad de que la prueba vaya a rechazar el contraste de hipótesis. Puesto que la probabilidad de un error de tipo II es la probabilidad de aceptar la hipótesis nula cuando la hipótesis alternativa es cierta, la potencia se puede expresar como (1-probabilidad de un error de tipo II), que es la probabilidad de rechazar la hipótesis nula cuando la hipótesis alternativa es cierta.

Estadísticos y gráficos. Prueba unilateral, prueba bilateral, nivel de significación, índice de errores de Tipo I, supuestos de prueba, desviación estándar de población, media de poblaciones bajo prueba, valor hipotetizado, potencia bidimensional por tamaño de muestra, potencia bidimensional por tamaño de efecto, potencia tridimensional por tamaño de muestra, potencia tridimensional por tamaño de efecto, grados de rotación, pares de grupos, coeficiente de correlación de momento-producto Pearson, diferencia de medias, rango de gráfico del tamaño de efecto, desviación estándar de población combinada, contrastes y diferencias por parejas, coeficientes de contraste, prueba de contraste, BONFERRONI, SIDAK, LSD, potencia por tamaño total de muestra, potencia bidimensional por desviación estándar combinada, potencia tridimensional por muestra total, potencia tridimensional por tamaño total de muestra, desviación estándar combinada, distribución T de Student, distribución T no central,

Consideraciones sobre datos de Power Analysis

Datos

En el análisis de una muestra, los datos observados se recopilan como una única muestra aleatoria.

En el análisis de muestras emparejadas, los datos observados contienen dos muestras emparejadas y correlacionadas y cada caso tiene dos mediciones.

En el análisis de muestras independientes, los datos observados contienen dos muestras independientes.

El análisis de varianza (ANOVA) es un método estadístico para realizar una estimación de las medias de varias poblaciones que, a menudo, se supone que se distribuyen normalmente.

Supuestos

En los análisis de una muestra, se supone que los datos de muestras siguen de forma independiente e idéntica una distribución normal con una media y una varianza fijas y extraen una inferencia estadística sobre el parámetro de media.

En los análisis de muestras emparejadas, se supone que los datos de cada muestra siguen de forma independiente e idéntica una distribución normal con una media y una varianza fijas y extraen una inferencia estadística sobre la diferencia de las dos medias.

En los análisis de muestras independientes, se supone que los datos de cada muestra siguen de forma independiente e idéntica una distribución normal con una media y una varianza fijas y extraen una inferencia estadística sobre la diferencia de las dos medias.

En el ANOVA de un factor, el método estadístico para realizar una estimación de las medias de varias poblaciones que, a menudo, se se supone que se distribuyen normalmente.

Cómo obtener un Power Analysis

1. En los menús seleccione:

   Analizar > **Power Analysis** > **Comparar medias** > **Prueba T de una muestra**, o **Prueba T de muestras emparejadas** o **Prueba T de muestra independiente** o **ANOVA de un factor**
2. Defina los supuestos de prueba necesarios.
3. Pulse **Aceptar**.

Medias

Las siguientes características de estadísticas se incluyen en IBM SPSS Estadística Base Edition.

Análisis de potencia de prueba T de una muestra

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

En el análisis de una muestra, los datos observados se recopilan como una única muestra aleatoria. Se supone que los datos de la muestra siguen de forma independiente e idéntica una distribución normal con una media y una varianza fija y extraen una inferencia estadística sobre el parámetro de la media.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Medias** > **Prueba T de una muestra**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores de potencia de cuadrícula** y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

   Puede obtener información adicional consultando "Análisis de potencia: Valores de cuadrícula" en la página 39.
4. Cuando se selecciona **Potencia** como método de **Estimación** de la asunción de pruebas, especifique el valor **Tamaño de muestra** apropiado para el valor de estimación de potencia. El valor debe ser un entero mayor que 1.
5. Opcionalmente, seleccione una opción del campo **Especificar**.

   Valores hipotéticos
   El valor predeterminado proporciona los valores **Media de población** y **Valor nulo**.

   Media de población
   Especifique un valor que especifique la media de población en pruebas. El valor debe ser un único valor numérico.

   Valor nulo
   Opcionalmente, especifique un valor que especifique el valor de hipótesis nulo que se va a probar. El valor debe ser un único valor numérico.

   Tamaño de efecto
   La f de Cohen se utiliza para estimar el tamaño de efecto como una entrada a la estimación de la potencia o el tamaño de la muestra. El tamaño de efecto definido **Valor** se pasa al paso intermedio del procedimiento y calcula la potencia o el tamaño de muestra deseado.
6. Introduzca un valor de **Desviación estándar de población**. El valor debe ser un solo número mayor que 0.
7. Seleccione si la prueba es unilateral o bilateral.

   Análisis no direccional (bilateral)
   Cuando está seleccionado, se utiliza una prueba bilateral. Ésta es el ajuste predeterminado.

   Análisis direccional (unilateral)
   Cuando está seleccionado, la potencia se calcula para una prueba unilateral.
8. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
9. Opcionalmente, puede pulsar **Gráfico** para especificar los valores de "Análisis de potencia de prueba T de una muestra: Gráfico" en la página 3 (salida de gráfica, valores de gráfico en dos dimensiones y valores de gráfico en tres dimensiones).

   Nota: Gráfico solo está disponible cuando se selecciona **Potencia** como supuesto de prueba.
10. De forma opcional, pulse **Precisión** para calcular el tamaño de muestra en función de los intervalos de confianza especificando los valores de la mitad del intervalo de confianza. Para obtener más información, consulte "Análisis de potencia: Precisión" en la página 38.

Nota: Precisión solo está disponible cuando se selecciona **Tamaño de muestra** como método **Estimación** de supuesto de pruebas, se selecciona **Valores hipotéticos** en la lista **Especificar** y se selecciona el análisis **No direccional (bilateral)** como **Dirección de prueba**.

Análisis de potencia de prueba T de una muestra: Gráfico

El diálogo **Gráfico** proporciona opciones para controlar los gráficos de salida para ilustrar la potencia de dos y tres dimensiones por medio de gráficos de tamaño de efecto o de muestra. El diálogo también controla los grados de rotación vertical y horizontal para gráficas tridimensionales.

Gráfico de dos dimensiones

Estimación de potencia versus tamaño de la muestra
Cuando está habilitado, este valor opcional proporciona opciones para controlar la potencia bidimensional mediante el gráfico del tamaño de la muestra. De forma predeterminada, el valor está inhabilitado.

Rango de tamaño de la muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de la muestra.

Límite inferior
Controla el límite inferior para la potencia bidimensional mediante el gráfico del tamaño de la muestra. El valor debe ser mayor que 1 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia bidimensional mediante el gráfico del tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Estimación de potencia versus tamaño del efecto
De forma predeterminada, este valor opcional está inhabilitado. Cuando está habilitado, el gráfico se muestra en los resultados. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de efecto.

Rango de tamaño del efecto
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles.

Límite inferior
Controla el límite inferior para la potencia bidimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor o igual que -5,0 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia bidimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor que el valor del **Límite inferior** y no puede ser mayor que 5,0.

Gráfico tridimensional

Estimación de potencia versus
Proporciona opciones para controlar la potencia tridimensional mediante el gráfico del tamaño de la muestra (eje x) y el tamaño de efecto (eje y), los valores de rotación vertical/horizontal y el rango de gráfico del tamaño de muestra/efecto especificado por el usuario. Este valor está inhabilitado de forma predeterminada.

Tamaño del efecto en el eje x y tamaño de la muestra en el eje y
El valor opcional controla la potencia tridimensional mediante el gráfico de tamaño de la muestra (eje x) y el tamaño de efecto (eje y). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Tamaño del efecto en el eje y y tamaño de la muestra en el eje x
El valor opcional controla la potencia tridimensional mediante el gráfico del tamaño de la muestra (eje y) y el tamaño de efecto (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Rango de tamaño de la muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de la muestra.

Límite inferior
Controla el límite inferior para la potencia tridimensional mediante el gráfico del tamaño de la muestra. El valor debe ser mayor que 1 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia tridimensional mediante el gráfico del tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Rango de tamaño del efecto
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de efecto.

Límite inferior
Controla el límite inferior para la potencia tridimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor o igual que -5,0 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia tridimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor que el valor del **Límite inferior** y no puede ser mayor que 5,0.

Análisis de potencia de prueba T para muestras emparejadas

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

En el análisis de muestras emparejadas, los datos observados contienen dos muestras emparejadas y correlacionadas y cada caso tiene dos mediciones. Se supone que los datos de cada muestra siguen de forma independiente e idéntica una distribución normal con una media y una varianza fija, y extraen una inferencia estadística sobre la diferencia de las medias.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Medias** > **Prueba T de muestras emparejadas**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores**

de potencia de cuadrícula y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

Puede obtener información adicional consultando "Análisis de potencia: Valores de cuadrícula" en la página 39.

4. Cuando se selecciona **Potencia** como método de **Estimación** de la asunción de pruebas, especifique el valor **Tamaño de muestra**. El valor debe ser un entero positivo mayor o igual que 2.
5. Opcionalmente, seleccione una opción del campo **Especificar**.

   Valores hipotéticos
   El valor predeterminado proporciona los valores **Media** y **Desviación estándar**.

   Diferencia media de la población
   Cuando es necesaria una única media de población, especifique un valor de **Diferencias de medias de población**. Cuando se especifica un único valor, denota la diferencia de medias de población μ_d.

   Nota: El valor no puede ser 0 cuando se selecciona **Tamaño de muestra**.

   Media de población del grupo 1 y del grupo 2
   Cuando se necesitan varias medias de población para los pares de grupos especificados, especifique valores para **Media de población del grupo 1** y **Media de población del grupo 2**. Cuando se especifican varios valores, denotan la diferencia de medias de la población μ_1 y μ_2.

   Nota: Los dos valores no pueden ser iguales cuando se selecciona **Tamaño de muestra** .

   Desviación estándar de la población para la diferencia de medias
   Cuando se especifica una única media de población, especifique la desviación estándar de población para el valor de diferencia de medias. Si se especifica un único valor, denota la desviación estándar de la población de la diferencia de grupos σ_d. El valor debe ser un solo número mayor que 0.

   Desviación estándar de la población para el grupo 1 y el grupo 2
   Cuando se especifican varias medias de la población, especifique la desviación estándar de la población para los valores del grupo 1 y del grupo 2. Cuando se especifican varios valores, denotan la desviación estándar de la población de la diferencia de grupos σ_1 y σ_2. Los valores deben ser números únicos mayores que 0.

   Coeficiente de correlación de momento-producto de Pearson
   Si lo desea, especifique un valor que especifique el coeficiente de correlación del momento-producto de Pearson ρ. El valor debe ser un único valor numérico entre -1 y 1. El valor no puede ser 0.

   Nota: Si se ha especificado un solo valor de **Desviación estándar de población para la diferencia de medias**, este valor se ignora. De lo contrario, los valores para **Desviación estándar de la población para el grupo 1 y el grupo 2** se utilizan para calcular σ_d.

   Tamaño de efecto
   Estima el tamaño del efecto como una entrada para la estimación de la potencia o el tamaño de la muestra. El tamaño de efecto definido **Valor** se pasa al paso intermedio del procedimiento y calcula la potencia o el tamaño de muestra deseado.
6. Seleccione si la prueba es unilateral o bilateral.

   Análisis no direccional (bilateral)
   Cuando está seleccionado, se utiliza una prueba bilateral. Ésta es el ajuste predeterminado.

   Análisis direccional (unilateral)
   Cuando está seleccionado, la potencia se calcula para una prueba unilateral.
7. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
8. Opcionalmente, pulse **Gráfico** para especificar los valores de "Análisis de potencia de prueba T para muestras emparejadas: Gráfico" en la página 7 (salida del gráfico, valores de gráfico bidimensionales, valores de gráfico tridimensional y ayudas contextuales).

Nota: Gráfico solo está disponible cuando se selecciona **Potencia** como supuesto de prueba.

9. De forma opcional, pulse **Precisión** para calcular el tamaño de muestra en función de los intervalos de confianza especificando los valores de la mitad del intervalo de confianza. Para obtener más información, consulte "Análisis de potencia: Precisión" en la página 38.

Nota: Precisión solo está disponible cuando se selecciona **Tamaño de muestra** como el método **Estimación** de asunción de pruebas y **Valores hipotéticos** está seleccionado en la lista **Especificar** .

Análisis de potencia de prueba T para muestras emparejadas: Gráfico

El diálogo **Gráfico** proporciona opciones para controlar los gráficos de salida para ilustrar la potencia de dos y tres dimensiones por medio de gráficos de tamaño de efecto o de muestra. El diálogo también controla los grados de rotación vertical y horizontal para gráficas tridimensionales.

Gráfico de dos dimensiones

Estimación de potencia versus tamaño de la muestra
Cuando está habilitado, este valor opcional proporciona opciones para controlar la potencia bidimensional mediante el gráfico del tamaño de la muestra. De forma predeterminada, el valor está inhabilitado.

Rango de tamaño de la muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de la muestra.

Límite inferior
Controla el límite inferior para la potencia bidimensional mediante el gráfico del tamaño de la muestra. El valor debe ser mayor que 1 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia bidimensional mediante el gráfico del tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Estimación de potencia versus tamaño del efecto
De forma predeterminada, este valor opcional está inhabilitado. Cuando está habilitado, el gráfico se muestra en los resultados. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de efecto.

Rango de tamaño del efecto
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles.

Límite inferior
Controla el límite inferior para la potencia bidimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor o igual que -5,0 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia bidimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor que el valor del **Límite inferior** y no puede ser mayor que 5,0.

Gráfico tridimensional

Estimación de potencia versus
Proporciona opciones para controlar la potencia tridimensional mediante el gráfico del tamaño de la muestra (eje x) y el tamaño de efecto (eje y), los valores de rotación vertical/horizontal y el rango de gráfico del tamaño de muestra/efecto especificado por el usuario. Este valor está inhabilitado de forma predeterminada.

Tamaño del efecto en el eje x y tamaño de la muestra en el eje y
El valor opcional controla la potencia tridimensional mediante el gráfico de tamaño de la muestra (eje x) y el tamaño de efecto (eje y). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Tamaño del efecto en el eje y y tamaño de la muestra en el eje x
El valor opcional controla la potencia tridimensional mediante el gráfico del tamaño de la muestra (eje y) y el tamaño de efecto (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Rango de tamaño de la muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de la muestra.

Límite inferior
Controla el límite inferior para la potencia tridimensional mediante el gráfico del tamaño de la muestra. El valor debe ser mayor que 1 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia tridimensional mediante el gráfico del tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Rango de tamaño del efecto
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de efecto.

Límite inferior
Controla el límite inferior para la potencia tridimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor o igual que -5,0 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia tridimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor que el valor del **Límite inferior** y no puede ser mayor que 5,0.

Análisis de potencia de prueba T para muestras independientes

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

En el análisis de muestras independientes, los datos observados contienen dos muestras independientes. Se supone que los datos de cada muestra siguen de forma independiente e idéntica una distribución normal con una media y una varianza fija, y extraen una inferencia estadística sobre la diferencia de las medias.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Medias** > **Prueba T de muestras independientes**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores de potencia de cuadrícula** y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

   Puede obtener información adicional consultando "Análisis de potencia: Valores de cuadrícula" en la página 39.

   Especifique un valor de **Proporción de tamaño de grupo** para especificar la proporción de los tamaños de las muestras (el valor debe ser un valor único entre 0,01 y 100.
4. Cuando se selecciona **Potencia** como método de **Estimación** de supuesto de pruebas, especifique valores para especificar el tamaño de muestra para **Tamaño de muestra para el grupo 1 y el grupo 2** para la comparación. Los valores deben ser enteros mayores que 1.
5. Opcionalmente, seleccione una opción del campo **Especificar**.

   Valores hipotéticos
   El valor predeterminado proporciona los valores **Media de población** y **Desviación estándar de población**.

   Diferencia media de la población
   Cuando se especifica una única media de población, especifique el valor de diferencia de media de población. Cuando se especifica un solo valor, indica la población de la diferencia de medias de grupo σ_d. El valor debe ser un solo número mayor que 0.

   Media de población del grupo 1 y del grupo 2
   Cuando se especifican varias medias de la población, especifique la media de población para los valores del grupo 1 y del grupo 2. Cuando se especifican varios valores, denotan la diferencia de medias de la población σ_1 y σ_2. Los valores deben ser números únicos mayores que 0.

   Las desviaciones estándar de población son
   Especifique si las desviaciones estándar de población son **Iguales para dos grupos** o **No iguales para dos grupos**.

   - Cuando las desviaciones estándar de población son iguales para dos grupos, especifique un valor para la **Desviación estándar combinada** que denote σ, y que suponga que las dos varianzas de grupo son iguales, o $\sigma_1 = \sigma_2 = \sigma$.
   - Cuando las desviaciones estándar de población no son iguales para dos grupos, especifique los valores para **Desviación estándar para el grupo 1 y el grupo 2** que denotan σ_1 y σ_2.

     Nota: Cuando los valores de **Desviación estándar para el grupo 1 y el grupo 2** son idénticos, se tratan como un único valor.

   Tamaño de efecto
   Estima el tamaño del efecto como una entrada para la estimación de la potencia o el tamaño de la muestra. El tamaño de efecto definido **Valor** se pasa al paso intermedio del procedimiento y calcula la potencia o el tamaño de muestra deseado.
6. Seleccione si la prueba es unilateral o bilateral.

   Análisis no direccional (bilateral)
   Cuando está seleccionado, se utiliza una prueba bilateral. Ésta es el ajuste predeterminado.

Análisis direccional (unilateral)
Cuando está seleccionado, la potencia se calcula para una prueba unilateral.

7. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
8. Opcionalmente, pulse **Gráfico** para especificar los valores de "Análisis de potencia de prueba T para muestras independientes: Gráfico" en la página 10 (salida del gráfico, valores de gráfico bidimensionales, valores de gráfico tridimensional y ayudas contextuales).

   Nota: Gráfico solo está disponible cuando se selecciona **Potencia** como supuesto de prueba.
9. De forma opcional, pulse **Precisión** para calcular el tamaño de muestra en función de los intervalos de confianza especificando los valores de la mitad del intervalo de confianza. Para obtener más información, consulte "Análisis de potencia: Precisión" en la página 38.

Nota: Precisión solo está disponible cuando se selecciona **Tamaño de muestra** como método **Estimación** de supuesto de pruebas, se selecciona **Valores hipotéticos** en la lista **Especificar** y se selecciona el análisis **No direccional (bilateral)** como **Dirección de prueba**.

Análisis de potencia de prueba T para muestras independientes: Gráfico

El diálogo **Gráfico** proporciona opciones para controlar los gráficos de salida para ilustrar la potencia de dos y tres dimensiones por medio de gráficas de proporción de muestra, tamaño de efecto o diferencia de medias. El diálogo también controla los grados de rotación vertical y horizontal para gráficas tridimensionales.

Gráfico de dos dimensiones

Estimación de potencia versus razón de tamaño de muestra/Estimación de potencia versus tamaño de muestra
Cuando está habilitado, este valor opcional proporciona opciones para controlar la potencia bidimensional mediante el gráfico de la proporción del tamaño de la muestra. Los valores adicionales no están disponibles cuando se han especificado varios valores de potencia (**Estimación de potencia versus tamaño de muestra**). De forma predeterminada, el valor está inhabilitado.

Rango de razón de tamaño de la muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de la muestra.

Límite inferior
Controla el límite inferior para la potencia bidimensional mediante el gráfico de proporción del tamaño de la muestra. El valor debe estar entre 0,01 y 100 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia bidimensional mediante el gráfico de proporción del tamaño de la muestra. El valor debe estar entre 0,01 y 100 y debe ser mayor que el valor de **Límite inferior**.

Estimación de potencia versus tamaño del efecto (o diferencia de medias)
De forma predeterminada, este valor opcional está inhabilitado. Cuando está habilitado, el gráfico se muestra en los resultados. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de efecto (o la diferencia de medias).

Rango de tamaño del efecto (o diferencia de medias)
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles.

Límite inferior
Controla el límite inferior para la potencia bidimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor o igual que -5,0 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia bidimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor que el valor del **Límite inferior** y no puede ser mayor que 5,0.

Gráfico tridimensional

Estimación de potencia versus
Proporciona opciones para controlar la potencia tridimensional mediante el gráfico de la proporción del tamaño de la muestra (eje x) y el tamaño de efecto (eje y), los valores de rotación vertical/horizontal y el rango de gráfico del tamaño de muestra/efecto especificado por el usuario. Este valor está inhabilitado de forma predeterminada.

Tamaño de efecto (o diferencia de medias) en el eje x y el tamaño de la muestra en el eje y.
El valor opcional controla la potencia tridimensional mediante el gráfico de la proporción del tamaño de la muestra (eje x) y el tamaño de efecto (eje y). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Tamaño de efecto (o diferencia de medias) en el eje y y tamaño de la muestra en el eje x.
El valor opcional controla la potencia tridimensional mediante el gráfico del tamaño de la muestra (eje y) y el tamaño de efecto (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Rango de razón de tamaño de la muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de la muestra.

Límite inferior
Controla el límite inferior para la potencia tridimensional mediante el gráfico de la proporción del tamaño de la muestra. El valor debe estar entre 0,01 y 100 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia tridimensional mediante el gráfico de la proporción del tamaño de la muestra. El valor debe estar entre 0,01 y 100 y debe ser mayor que el valor de **Límite inferior**.

Rango de tamaño del efecto (o diferencia de medias)
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de efecto.

Límite inferior
Controla el límite inferior para la potencia tridimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor o igual que -5,0 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para la potencia tridimensional mediante el gráfico del tamaño de efecto. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Power Analysis del ANOVA de un factor

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

El análisis de varianza (ANOVA) es un método estadístico para realizar una estimación de las medias de varias poblaciones que, a menudo, se supone que se distribuyen normalmente. El ANOVA de un factor, un tipo común de ANOVA, es una extensión de la prueba *t* de dos muestras. El procedimiento proporciona aproximaciones para realizar una estimación de la potencia para los dos tipos de hipótesis para comparar las distintas medias de grupos, la prueba general y la prueba con contrastes especificados. La prueba general se centra en la hipótesis nula de que todas las medias de grupos son iguales. La prueba con contrastes especificados desglosa las hipótesis de ANOVA generales en partes más pequeñas, pero que son más descriptibles y útiles de los medios.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Medias** > **ANOVA de un factor**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores de potencia de cuadrícula** y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

   Puede obtener información adicional consultando “Análisis de potencia: Valores de cuadrícula” en la página 39.
4. Opcionalmente, seleccione una opción del campo **Especificar**.

   Valores hipotéticos
   El valor predeterminado proporciona el valor **Desviación estándar de población combinada**. Especifique un único valor numérico mayor que 0. El valor predeterminado es 1.

   Tamaño de efecto: f de Cohen
   La f de Cohen se utiliza para estimar el tamaño de efecto como una entrada a la estimación de la potencia o el tamaño de la muestra. El tamaño de efecto definido **Valor** se pasa al paso intermedio del procedimiento y calcula la potencia o el tamaño de muestra deseado.

   Tamaño de efecto: eta cuadrado
   Eta-cuadrado (η^2) se utiliza para estimar el tamaño de efecto como una entrada a la estimación de la potencia o del tamaño de la muestra. El tamaño de efecto definido **Valor** se pasa al paso intermedio del procedimiento y calcula la potencia o el tamaño de muestra deseado.
5. Cuando se selecciona **Potencia** como valor de **Estimación** de supuesto de pruebas, especifique los valores **Tamaños de grupo** y **Medias de grupo**. Se deben especificar al menos dos valores de tamaño de grupo, y cada uno de los valores de tamaño de grupo debe ser mayor o igual a 2 (pero no puede superar 2,147, 483, 647). Se debe especificar, al menos, dos valores de medias de grupos (el número de valores especificados debe ser igual a los valores de tamaño de grupo).

   Cuando se selecciona **Tamaño de muestra** como valor de **Estimación** de supuesto de pruebas, especifique los valores **Ponderaciones de grupo** (opcional) y **Medias de grupo**. Las ponderaciones de grupo asignan las ponderaciones de tamaño de grupo cuando se selecciona **Potencia**.

   Nota: Los valores de **Ponderaciones de grupo** se ignoran cuando se especifican los valores **Tamaños de grupo**.

Pulse **Añadir** para incluir más valores de **Tamaños de grupo**, **Ponderaciones de grupo** y **Medias de grupo**. Pulse **Suprimir** para eliminar los valores existentes.

6. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
7. De forma opcional, pulse **Contraste** para especificar los valores de "Análisis de potencia de ANOVA de un factor: Contraste" en la página 13 (prueba de contraste, diferencias por parejas, tamaños de efecto y estimación del tamaño de muestra en función de los intervalos de confianza).
8. Opcionalmente, pulse **Gráfico** para especificar los valores de "Análisis de potencia de ANOVA de un factor: Gráficos" en la página 14 (salida del gráfico, valores de gráfico bidimensionales, valores de gráfico tridimensional y ayudas contextuales).

   Nota: Gráfico solo está disponible cuando se especifican valores de **Tamaños de grupo** y se selecciona **Potencia** como valor de **Estimación**.
9. De forma opcional, pulse **Precisión** para calcular el tamaño de muestra en función de los intervalos de confianza especificando los valores de la mitad del intervalo de confianza. Para obtener más información, consulte "Análisis de potencia: Precisión" en la página 38.

Nota: Precisión solo está disponible cuando se selecciona **Tamaño de muestra** como método de **Estimación** de supuesto de pruebas.

Análisis de potencia de ANOVA de un factor: Contraste

El diálogo **Contraste** proporciona opciones para especificar el contraste, el coeficiente, el tamaño de efecto, las diferencias por parejas y los valores de la mitad del ancho del intervalo de confianza para el análisis de potencia de análisis de ANOVA de un factor.

Prueba de contraste

Prueba con contraste lineales
Cuando está habilitado, están disponibles los valores de contraste y coeficiente.

Dirección de la prueba

Análisis no direccional (bilateral)
Cuando está seleccionado, se utiliza una prueba bilateral. Ésta es el ajuste predeterminado.

Análisis direccional (unilateral)
Cuando está seleccionado, la potencia se calcula para una prueba unilateral.

Coeficientes
Utilice la tabla para especificar los coeficientes de contraste y solicitar la prueba de contraste. Los valores de tabla son opcionales. El número de valores especificados debe ser igual a los valores especificado para **Tamaños de grupo** y **Medias de grupo**. La suma de todos los valores especificados debe ser igual a 0, de lo contrario, el último valor se ajustará automáticamente.

Tamaño de efecto
Especifique el tamaño de efecto de la prueba de contraste que mide δ. Debe especificar un único valor numérico. Cuando **Dirección de ensayo** se establece en **Análisis direccional (unilateral)**, el valor especificado debe ser ≥ 0. Cuando se especifica **Potencia** como el supuesto de prueba de **Estimación**, el valor especificado no puede ser 0.

Diferencias por parejas

Estimar la potencia de prueba para diferencias por parejas
Controla si se va a realizar una estimación de la potencia de las pruebas para las diferencias por parejas. De forma predeterminada, el valor opcional está inhabilitado, lo que suprime los resultados para las diferencias por parejas.

Corrección de Bonferroni
Utilice la corrección de Bonferroni en la estimación de la potencia de las diferencias por parejas. Ésta es el ajuste predeterminado.

corrección de Sidak
Utiliza la corrección de Sidak en la estimación de la potencia de las diferencias por parejas.

Diferencia menos significativa (LSD)
Utiliza la corrección de diferencia menos significativa (LSD) en la estimación de la potencia de diferencias por parejas.

Especifique la mitad del ancho del intervalo de confianza
Estima el tamaño de muestra basándose en el valor de medio ancho del intervalo de confianza. Especifique un valor en el rango de 0 a 1. Para la prueba binomial de una muestra, el valor debe estar en el rango de 0 a 0,5.

Nota: Se ignoran los valores duplicados.

- Pulse **Añadir** para añadir el valor de media anchura especificado a la lista.
- Resalte un valor de media anchura existente y pulse **Cambiar** para actualizar el valor.
- Resalte un valor de media anchura existente y pulse **Eliminar** para eliminar el valor de la lista.

Análisis de potencia de ANOVA de un factor: Gráficos

El diálogo **Gráfico** controla los gráficos de salida para ilustrar los gráficos de tamaño de efecto y muestra de potencia bidimensional y tridimensional. El diálogo también controla los grados de rotación vertical y horizontal para gráficas tridimensionales.

Gráfico de dos dimensiones

Estimación de potencia versus tamaño total de la muestra
Cuando está habilitado, este valor opcional proporciona opciones para controlar la potencia bidimensional mediante el gráfico del tamaño total de la muestra. De forma predeterminada, el valor está inhabilitado.

Rango de tamaño total de la muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño total de la muestra.

Límite inferior
Controla el límite inferior para la potencia bidimensional mediante el gráfico de tamaño total de la muestra. El valor debe ser mayor o igual que:

- 2 x el número de enteros especificados para **Tamaños de grupo**
- 2 x la suma de los enteros especificados para **Tamaños de grupo** / por el valor de entero menor para los **Tamaños de grupo**

El valor no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la potencia bidimensional mediante el gráfico del tamaño total de la muestra. El valor debe ser menor o igual que.

- 5000 / por el valor de entero superior especificado para los **Tamaños de grupo** x la suma de los enteros especificados para los **Tamaños de grupo**

El valor debe ser mayor que el **Límite inferior** y no puede ser mayor que 2147483647.

Estimación de potencia versus desviación estándar combinada
De forma predeterminada, este valor opcional está inhabilitado. El valor controla la potencia bidimensional mediante el gráfico de desviación estándar. Cuando está habilitado, el gráfico se muestra en los resultados. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado de la desviación estándar combinada.

Nota:

El gráfico está inhabilitado cuando los valores especificados de **Medias de grupo** son todos iguales.

Rango de desviación estándar combinada
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles.

Límite inferior
Controla el límite inferior para la potencia bidimensional mediante el gráfico de desviación estándar combinada. El valor debe ser mayor que 0 y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la potencia bidimensional mediante el gráfico de la desviación estándar combinada. El valor debe ser mayor que el valor del **Límite inferior**.

Gráfico tridimensional

Estimación de potencia versus
Proporciona opciones para controlar la potencia tridimensional por tamaño de muestra total (eje x) y tamaño de efecto (eje y), los valores de rotación vertical y horizontal y el rango de gráfico especificado por el usuario de tamaño de muestra y efecto. Este valor está inhabilitado de forma predeterminada.

Nota:

El gráfico está inhabilitado cuando los valores especificados de **Medias de grupo** son todos iguales.

Desviación estándar combinada en el eje x y tamaño total de la muestra en el eje y
El valor opcional controla la potencia tridimensional mediante un gráfico de tamaño total de la muestra (eje x) y de desviación estándar combinada (eje y). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Desviación estándar combinada en el eje y y tamaño total de la muestra en el eje x
El valor opcional controla la potencia tridimensional mediante el gráfico de tamaño total de la muestra (eje y) y la desviación estándar combinada (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Rango de tamaño total de la muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de la muestra.

Límite inferior
Controla el límite inferior para la potencia tridimensional mediante el gráfico del tamaño total de la muestra. El valor debe ser mayor que 0 y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la potencia tridimensional mediante el gráfico del tamaño total de la muestra. El valor debe ser mayor que el valor del **Límite inferior**.

Rango de desviación estándar combinada
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de efecto.

Límite inferior
Controla el límite inferior para la potencia tridimensional mediante el gráfico de la desviación estándar combinada. El valor debe ser mayor que 0 y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la potencia tridimensional mediante el gráfico de la desviación estándar combinada. El valor debe ser mayor que el valor del **Límite inferior**.

Estimación de potencia versus tamaño de la muestra
Proporciona opciones para controlar la potencia tridimensional por tamaño de efecto y el rango de gráfico especificado por el usuario de tamaño de muestra y efecto. Este valor está inhabilitado de forma predeterminada.

Nota:

El gráfico está inhabilitado cuando los valores especificados de **Medias de grupo** son todos iguales.

Desviación estándar combinada en el eje x y tamaño total de la muestra en el eje y
El valor opcional controla la potencia tridimensional por tamaño de muestra total (eje x) y gráfico de tamaño de efecto (eje y). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Desviación estándar combinada en el eje y y tamaño total de la muestra en el eje x
El valor opcional controla la potencia tridimensional por tamaño total de muestra (eje y) y gráfico de tamaño de efecto (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Rango de tamaño total de la muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de la muestra.

Límite inferior
Controla el límite inferior para la potencia tridimensional mediante el gráfico del tamaño total de la muestra. El valor debe ser mayor que 0 y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la potencia tridimensional mediante el gráfico del tamaño total de la muestra. El valor debe ser mayor que el valor del **Límite inferior**.

Rango de tamaño del efecto
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifica ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado del tamaño de efecto.

Límite inferior
Controla el límite inferior para la potencia tridimensional mediante el gráfico del tamaño total de la muestra. El valor debe ser mayor que 0 y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la potencia tridimensional mediante el gráfico del tamaño total de la muestra. El valor debe ser mayor que el valor del **Límite inferior**.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Proporciones

Las siguientes características de estadísticas se incluyen en IBM SPSS Estadística Base Edition.

Análisis de potencia de prueba binominal para una muestra

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

La distribución binomial se basa en una secuencia de pruebas de Bernoulli. Se puede utilizar para modelar experimentos, incluyendo un número fijo de ensayos totales, que se supone que son independientes entre sí. Cada ensayo lleva a un resultado dicotómico, con la misma probabilidad para un resultado satisfactorio.

La prueba binomial para una muestra realiza la inferencia estadística sobre el parámetro de proporción comparándolo con un valor hipotetizado. Los métodos para estimar la potencia de una prueba de este calibre son la aproximación normal o la enumeración binomial.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Proporciones** > **Prueba binomial de una muestra**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores de potencia de cuadrícula** y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

   Puede obtener información adicional consultando “Análisis de potencia: Valores de cuadrícula” en la página 39.
4. Cuando se selecciona **Potencia** como valor de **Estimación** de supuesto de pruebas, especifique el valor correspondiente de **Número total de ensayos**. El valor debe ser un entero mayor o igual que 1.
5. Especifique un valor que especifique el valor de hipótesis alternativo del parámetro de proporción en el campo **Proporción de población**. El valor debe ser un único valor numérico.

   Nota: Cuando se especifica un valor de **Potencia**, el valor de **Proporción de población** no puede ser igual al **Valor nulo**.
6. Opcionalmente, entre un valor que especifique el valor de hipótesis nula del parámetro de proporción que se va a probar en el campo **Valor nulo**. El valor debe ser un único valor numérico en el rango 0 - 1. El valor predeterminado es 0.50.
7. Seleccione un método para estimar la potencia.

   Aproximación normal
   Habilita aproximación normal. Ésta es el ajuste predeterminado.

   Aplicar corrección de continuidad
   Controle si se utiliza o no la corrección de continuidad para el método de aproximación normal.

   Enumeración binomial
   Habilita el método de enumeración binomial. Opcionalmente, utilice el campo **Límite de tiempo** para especificar el número máximo de minutos permitido para estimar el tamaño de la muestra. Cuando se alcanza el límite de tiempo, se termina el análisis y se muestra un mensaje de advertencia. Cuando se especifica, el valor debe ser un entero positivo único para indicar el número de minutos. El valor predeterminado es 5 minutos.

Nota: El supuesto de estimación de potencia seleccionada no tiene ningún efecto cuando el valor de **Número total de ensayos** supera los 500.

8. Seleccione si la prueba es unilateral o bilateral.

   Análisis no direccional (bilateral)
   Cuando está seleccionado, se utiliza una prueba bilateral. Ésta es el ajuste predeterminado.

   Análisis direccional (unilateral)
   Cuando está seleccionado, la potencia se calcula para una prueba unilateral.

9. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
10. Opcionalmente, pulse **Gráfico** para especificar los valores de "Análisis de potencia de binominal para una muestra: Gráfico" en la página 18 (salida de gráfica, valores de gráfico en dos dimensiones y valores de gráfico en tres dimensiones).

    Nota: Gráfico solo está disponible cuando se selecciona **Potencia** como supuesto de prueba de **Estimación** y **Enumeración binomial** no está seleccionado.

11. De forma opcional, pulse **Precisión** para calcular el tamaño de muestra en función de los intervalos de confianza especificando los valores de la mitad del intervalo de confianza. Para obtener más información, consulte "Análisis de potencia: Precisión" en la página 38.

Nota: Precisión solo está disponible cuando se selecciona **Tamaño de muestra** como método de **Estimación** de supuesto de prueba y se selecciona el análisis **No direccional (bilateral)** como **Dirección de prueba**.

Análisis de potencia de binominal para una muestra: Gráfico

El diálogo **Gráfico** proporciona opciones para controlar los gráficos de salida para ilustrar la potencia bidimensional y tridimensional mediante gráficas. El diálogo también controla los grados de rotación vertical y horizontal para gráficas tridimensionales.

Gráfico de dos dimensiones
Proporciona opciones para controlar la estimación de potencia bidimensional versus los gráficos. Este valor está inhabilitado de forma predeterminada.

Estimación de potencia versus valor de hipótesis nula
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante un gráfico de valores nulos. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus valor de hipótesis alternativa
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante un gráfico de valores alternativos. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus la diferencia entre valores hipotetizados
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de diferencia entre valores hipotetizados. De forma predeterminada, el valor está inhabilitado.

Estimación de potencia versus número total de ensayos
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el número total de ensayos de gráficos. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Rango de gráfico de número total de ensayos
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para el gráfico de potencia bidimensional versus el número total de ensayos de gráficos. El valor debe ser mayor que 0 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para el gráfico de potencia bidimensional versus el número total de ensayos de gráficos. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Gráfico tridimensional
Proporciona opciones para controlar la estimación de potencia tridimensional versus gráficos. Este valor está inhabilitado de forma predeterminada.

Estimación de potencia versus número total de ensayos
Cuando se selecciona, este valor habilita las opciones siguientes.

en el eje x y la diferencia entre los valores hipotetizados en el eje y
El valor opcional controla la potencia tridimensional mediante un gráfico de número total de ensayos (eje x) y del gráfico de diferencia entre valores hipotetizados (eje y). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

en el eje y y la diferencia entre los valores hipotetizados en el eje x
El valor opcional controla la potencia tridimensional mediante el número total de ensayos (eje y) y el gráfico de diferencia entre valores hipotetizados (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Rango de gráfico de número total de ensayos
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para el gráfico de potencia tridimensional versus el número total de ensayos de gráficos. El valor debe ser mayor que 0 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para el gráfico de potencia tridimensional versus el número total de ensayos de gráficos. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Estimación de potencia versus valor de hipótesis nula
Cuando se selecciona, este valor habilita las opciones siguientes.

en el eje x y el valor de hipótesis alternativa en el eje y
El valor opcional controla la potencia tridimensional mediante un gráfico de valor alternativo y nulo (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

en el eje y y el valor de hipótesis alternativa en eje x
El valor opcional controla la potencia tridimensional mediante un gráfico de valor nulo y alternativo (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Análisis de potencia de prueba binomial para muestras relacionadas

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

La distribución binomial se basa en una secuencia de pruebas de Bernoulli. Se puede utilizar para modelar experimentos, incluyendo un número fijo de ensayos totales, que se supone que son independientes entre sí. Cada ensayo lleva a un resultado dicotómico, con la misma probabilidad para un resultado satisfactorio.

La prueba binominal para muestras relacionadas estima la potencia de prueba de McNemar para comparar dos parámetros de proporción basados en los sujetos de pares coincidentes de dos poblaciones binominales relacionadas.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Proporciones** > **Prueba binomial de muestra relacionada**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores de potencia de cuadrícula** y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

   Puede obtener información adicional consultando “Análisis de potencia: Valores de cuadrícula” en la página 39.
4. Cuando se selecciona **Potencia** como método de **Estimación** de supuesto de pruebas, especifique el valor correspondiente de **Número total de pares**.
5. Seleccione esta opción para especificar valores de prueba para **Proporciones** o **Recuentos**.
   - Cuando se selecciona **Proporciones**, escriba los valores en los campos **Proporción 1** y **Proporción 2**. Los valores deben estar en el rango 0 - 1.
   - Cuando se selecciona **Recuentos**, especifique los valores en los campos **Recuento 1** y **Recuento 2**. Los valores deben estar entre 0 y el valor especificado para **Número total de pares**.

   Notas sobre proporciones:

   - **Proporciones** es la única opción disponible cuando se especifica un valor de **Potencia**.
   - Cuando no se selecciona **Los valores de prueba son marginales**, 0 < **Proporción 1** + **Proporción 2** ≤ 1
   - Cuando **Los valores de prueba son marginales** está seleccionado:
     - **Proporción 1** * **Proporción 2** > 0
     - **Proporción 1** < 1
     - **Proporción 2** < 1
     - Los valores para **Proporción 1** y **Proporción 2** no pueden ser los mismos.

   Notas sobre recuentos:

   - Los valores de **recuento** solo están disponibles cuando se selecciona **Potencia** como valor de **Estimación** de supuesto de pruebas.
   - Cuando **Los valores de prueba son marginales** no está seleccionado: 0 < **Recuento 1** + **Recuento 2** ≤ **Número total de pares**
   - Cuando **Los valores de prueba son marginales** está seleccionado:

- **Recuento 1** * **Recuento 2** > 0
- **Recuento 1** < **Número total de pares**
- **Recuento 2** < **Número total de pares**

6. Puede seleccionar opcionalmente **Los valores de prueba son marginales** para controlar si las proporciones especificadas o los valores de recuentos son marginales. Cuando **Los valores de prueba son marginales** está habilitado, debe especificar un valor **Correlación entre pares coincidentes**. El valor debe ser un valor único entre -1 y 1.
7. Seleccione un método para estimar la potencia.

   Aproximación normal
   Habilita aproximación normal. Ésta es el ajuste predeterminado.

   Enumeración binomial
   Habilita el método de enumeración binomial. Opcionalmente, utilice el campo **Límite de tiempo** para especificar el número máximo de minutos permitido para estimar el tamaño de la muestra. Cuando se alcanza el límite de tiempo, se termina el análisis y se muestra un mensaje de advertencia. Cuando se especifica, el valor debe ser un entero positivo único para indicar el número de minutos. El valor predeterminado es 5 minutos.
8. Seleccione si la prueba es unilateral o bilateral.

   Análisis no direccional (bilateral)
   Cuando está seleccionado, se utiliza una prueba bilateral. Ésta es el ajuste predeterminado.

   Análisis direccional (unilateral)
   Cuando está seleccionado, la potencia se calcula para una prueba unilateral.
9. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
10. Opcionalmente, pulse **Gráfico** para especificar los valores de "Análisis de potencia de prueba binomial para muestras relacionadas" en la página 19 (salida de gráfica, valores de gráfico en dos dimensiones y valores de gráfico en tres dimensiones).

    Nota: Gráfico solo está disponible cuando se selecciona **Potencia** como supuesto de prueba de **Estimación** y **Enumeración binomial** no está seleccionado.
11. De forma opcional, pulse **Precisión** para calcular el tamaño de muestra en función de los intervalos de confianza especificando los valores de la mitad del intervalo de confianza. Para obtener más información, consulte "Análisis de potencia: Precisión" en la página 38.

Nota: Precisión solo está disponible cuando se selecciona **Tamaño de muestra** como método de **Estimación** de supuesto de prueba y se selecciona el análisis **No direccional (bilateral)** como **Dirección de prueba**.

Análisis de potencia de prueba binomial para muestras relacionadas: Gráfico

El diálogo **Gráfico** controla los gráficos de salida para ilustrar las gráficas de estimación de potencia bidimensional y tridimensional. El diálogo también controla los grados de rotación vertical y horizontal para gráficas tridimensionales.

Gráfico de dos dimensiones
Proporciona opciones para controlar la estimación de potencia bidimensional versus los gráficos. Este valor está inhabilitado de forma predeterminada.

Estimación de potencia versus número total de pares
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el número total de pares de gráficos. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Rango de gráfico de número total de pares
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para el gráfico de potencia bidimensional versus el número total de pares de gráficos. El valor debe ser mayor que 1 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para el gráfico de potencia bidimensional versus el número total de pares de gráficos. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 2500.

Estimación de potencia versus diferencia de riesgo
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de diferencia de riesgo. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus proporción de riesgo
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de proporción de riesgo. De forma predeterminada, el valor está inhabilitado.

Rango de gráfico de proporción de disco
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional versus el gráfico de proporción de riesgo. El valor no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional versus el gráfico de proporción de riesgo. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 10.

Estimación de potencia versus razón de probabilidad
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de razón de probabilidad. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Rango de gráfico de razón de probabilidad
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional versus el gráfico de razón de probabilidad. El valor no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional versus el gráfico de razón de probabilidad. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 10.

Estimación de potencia versus correlación entre pares emparejados
Controla el valor estimado de potencia power Ω frente a la correlación entre el gráfico de pares coincidentes. El gráfico se crea solo cuando se especifican las proporciones o recuentos marginales (en lugar de las proporciones discordantes).

Gráfico tridimensional
Proporciona opciones para controlar la estimación de potencia tridimensional versus gráficos. Este valor está inhabilitado de forma predeterminada.

Estimación de potencia versus proporciones discordantes
Cuando está habilitado, este valor opcional controla la potencia tridimensional mediante el gráfico de proporciones discordantes. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus proporciones marginales
Cuando está habilitado, este valor opcional controla la potencia tridimensional mediante el gráfico de proporciones marginales. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Nota: Este valor sólo está disponible cuando se selecciona **Los valores de prueba son marginales**.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Análisis de potencia de prueba binomial de muestras independientes

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

La distribución binomial se basa en una secuencia de ensayos de Bernoulli. Se puede utilizar para modelar estos experimentos, incluyendo un número fijo de ensayos totales que se supone que son independientes entre sí. Cada ensayo lleva a un resultado dicotómico, con la misma probabilidad para un resultado "satisfactorio". La prueba binominal de muestras independientes compara dos parámetros de proporción independientes.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Proporciones** > **Prueba binomial de muestras independientes**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores de potencia de cuadrícula** y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

   Puede obtener información adicional consultando “Análisis de potencia: Valores de cuadrícula” en la página 39.

   Opcionalmente, especifique un valor de **Proporción de tamaño de grupo**. El valor predeterminado es 1.
4. Cuando se selecciona **Potencia** como valor de **Estimación** de supuesto de prueba, indique valores para especificar el **Número total de ensayos para el grupo 1 y el grupo 2**. Los valores deben ser enteros mayores que 1.
5. Especifique los parámetros de proporción para los dos grupos. Ambos valores deben estar en el rango 0 - 1.

   Nota: Los dos valores no pueden ser el mismo cuando se especifica un valor de **Potencia**.

6. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
7. Seleccione el método de prueba deseado:

   Prueba de chi cuadrado
   Estima la potencia basada en la prueba de chi-cuadrado de Pearson. Ésta es el ajuste predeterminado.

   La desviación estándar se ha agrupado
   Este valor opcional controla si la estimación de la desviación estándar se ha agrupado o desagrupado. Este valor está habilitado de forma predeterminada.

   Aplicar corrección de continuidad
   Este ajuste opcional controla si se utiliza o no la corrección de continuidad. De forma predeterminada, el valor está inhabilitado.

   Prueba T
   Estima la potencia basada en la prueba T de Student.

   La desviación estándar se ha agrupado
   Este valor opcional controla si la estimación de la desviación estándar se ha agrupado o desagrupado. Este valor está habilitado de forma predeterminada.

   Prueba de razón de verosimilitud
   Estima la potencia basada en la prueba de razón de verosimilitud.

   prueba exacta de Fisher
   Estima la potencia basada en la prueba exacta de Fisher.

   Notas:

   - En algunos casos, la prueba exacta de Fisher puede tardar un periodo prolongado de tiempo en completarse.
   - Todos los gráficos se bloquean cuando se selecciona la prueba exacta de Fisher.
8. Seleccione un método para estimar la potencia.

   Aproximación normal
   Habilita aproximación normal. Ésta es el ajuste predeterminado.

   Enumeración binomial
   Habilita el método de enumeración binomial. Opcionalmente, utilice el campo **Límite de tiempo** para especificar el número máximo de minutos permitido para estimar el tamaño de la muestra. Cuando se alcanza el límite de tiempo, se termina el análisis y se muestra un mensaje de advertencia. Cuando se especifica, el valor debe ser un entero positivo único para indicar el número de minutos. El valor predeterminado es 5 minutos.
9. Seleccione si la prueba es unilateral o bilateral.

   Análisis no direccional (bilateral)
   Cuando está seleccionado, se utiliza una prueba bilateral. Ésta es el ajuste predeterminado.

   Análisis direccional (unilateral)
   Cuando está seleccionado, la potencia se calcula para una prueba unilateral.
10. Opcionalmente, puede pulsar **Gráfico** para especificar los valores de "Análisis de potencia de prueba binomial para ejemplos independientes: Gráfico" en la página 25 (salida de gráfica, valores de gráfico en dos dimensiones y valores de gráfico en tres dimensiones).

    Nota: Gráfico solo está disponible cuando se selecciona **Potencia** como supuesto de prueba de **Estimación** y **Enumeración binomial** no está seleccionado.
11. De forma opcional, pulse **Precisión** para calcular el tamaño de muestra en función de los intervalos de confianza especificando los valores de la mitad del intervalo de confianza. Para obtener más información, consulte "Análisis de potencia: Precisión" en la página 38.

Nota: Precisión solo está disponible cuando se selecciona **Tamaño de muestra** como método de **Estimación** de supuesto de prueba y se selecciona el análisis **No direccional (bilateral)** como **Dirección de prueba**.

Análisis de potencia de prueba binomial para ejemplos independientes: Gráfico

El diálogo **Gráfico** proporciona opciones para controlar los gráficos de salida para ilustrar las gráficas de estimación de potencia bidimensional y tridimensional. El diálogo también controla los grados de rotación vertical y horizontal para gráficas tridimensionales.

Gráfico de dos dimensiones
Proporciona opciones para controlar la estimación de potencia bidimensional versus los gráficos. Este valor está inhabilitado de forma predeterminada.

Estimación de potencia versus razón de tamaño de grupo/Estimación de potencia versos tamaño de grupo
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de razón de tamaño del grupo. Cuando se selecciona, este valor muestra el gráfico. Los valores adicionales no están disponibles cuando se han especificado varios valores de potencia (**Estimación de potencia versus tamaño de grupo**). De forma predeterminada, el valor está inhabilitado.

Rango de gráfico de proporción de tamaño de grupo
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para el gráfico de potencia bidimensional versus el número total de pares de gráficos. El valor debe ser mayor que .01 y no puede ser mayor que el valor de **Límite superior**.

Límite superior
Controla el límite superior para el gráfico de potencia bidimensional versus el número total de pares de gráficos. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 100.

Estimación de potencia versus diferencia de riesgo
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de diferencia de riesgo. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus proporción de riesgo
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de proporción de riesgo. De forma predeterminada, el valor está inhabilitado.

Rango de gráfico de proporción de disco
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional versus el gráfico de proporción de riesgo. El valor no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional versus el gráfico de proporción de riesgo. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 10.

Estimación de potencia versus razón de probabilidad
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de razón de probabilidad. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Rango de gráfico de razón de probabilidad
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional versus el gráfico de razón de probabilidad. El valor no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional versus el gráfico de razón de probabilidad. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 10.

Gráfico tridimensional
Proporciona opciones para controlar la estimación de potencia tridimensional versus gráficos. Este valor está inhabilitado de forma predeterminada.

Estimación de potencia versus proporciones
Cuando se selecciona, este valor opcional proporciona las siguientes opciones de potencia mediante la proporción:

proporción del grupo 1 en el eje x y proporción del grupo 2 en el eje y
Controla la potencia tridimensional mediante la proporción del Grupo 1 (eje x) y el gráfico de proporción de Grupo 2 (eje y). De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

proporción del grupo 1 en el eje y y proporción del grupo 2 en el eje x
Controla la potencia tridimensional mediante la proporción del Grupo 2 (eje x) y el gráfico de proporción de Grupo 1 (eje y). De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus tamaños de grupo
Cuando se selecciona esta opción, este valor opcional proporciona las siguientes opciones de potencia por tamaños de grupo:

tamaño del grupo 1 en el eje x y tamaño del grupo 2 en el eje y
Controla la potencia tridimensional mediante el número de ensayos en el grupo 1 (eje x) y el número de ensayos en el gráfico del grupo 2 (eje y). De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

tamaño del grupo 1 en el eje y y tamaño del grupo 2 en el eje x
Controla la potencia tridimensional mediante el número de ensayos en el grupo 2 (eje x) y el número de ensayos en el gráfico del grupo 1 (eje y). De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Rango de gráfico de tamaño del grupo 1
Cuando se selecciona, las opciones de límite inferior y superior para el gráfico del grupo 1 están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional versus el gráfico de razón de probabilidad. El valor debe ser mayor o igual que 2, y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional versus el gráfico de razón de probabilidad. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 2500.

Rango de gráfico de tamaño del grupo 2
Cuando se selecciona, las opciones de límite inferior y superior para el gráfico del grupo 2 están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional versus el gráfico de razón de probabilidad. El valor debe ser mayor o igual que 2, y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional versus el gráfico de razón de probabilidad. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 2500.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Correlaciones

Las siguientes características de estadísticas se incluyen en IBM SPSS Estadística Base Edition.

Análisis de potencia de prueba de correlación de Pearson para una muestra

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

El coeficiente de correlación de Momento-producto de Pearson mide la intensidad de la asociación lineal entre dos variables aleatorias de escala que se presuponen que siguen una distribución normal bivariada. Por convención, es una cantidad sin dimensión y obtenida estandarizando la covarianza entre dos variables continuas, con lo que oscila entre -1 y 1.

La prueba utiliza el método asintótico de Fisher para estimar la potencia para la correlación de Pearson de una muestra.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Correlaciones** > **Momento-producto de Pearson**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores de potencia de cuadrícula** y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

   Puede obtener información adicional consultando “Análisis de potencia: Valores de cuadrícula” en la página 39.

4. Cuando se selecciona **Potencia** como valor de **Estimación** de supuesto de prueba, especifique el valor correspondiente de **Tamaño de muestra en pares**. El valor debe ser un entero único mayor que 3.
5. Especifique un valor que especifique el valor de hipótesis alternativo del parámetro de correlación en el campo **Parámetro de correlación de Pearson**. El valor debe ser un único numérico entre -1 y 1.

   Nota: Cuando se especifica **Potencia**, el valor de **Parámetro de correlación de Pearson** no puede ser -1 o 1 y no puede ser igual al **Valor nulo**.
6. Opcionalmente, entre un valor que especifique el valor de hipótesis nulo del parámetro de correlación que se probará en el campo **Valor nulo**. El valor debe ser un único numérico entre -1 y 1. El valor predeterminado es 0.

   Nota: Cuando se especifica **Potencia**, **Valor nulo** no puede ser -1 o 1.
7. Opcionalmente, seleccione **Utilizar fórmula de corrección de sesgo en la estimación de potencia** para especificar si el ajuste de sesgo está implicado o se ignora. El valor se habilita de forma predeterminada, lo que incluye el término de ajuste de sesgo en la estimación de potencia. Cuando el valor no está seleccionado, se ignora el término de ajuste de sesgo.
8. Seleccione si la prueba es unilateral o bilateral.

   Análisis no direccional (bilateral)
   Cuando está seleccionado, se utiliza una prueba bilateral. Ésta es el ajuste predeterminado.

   Análisis direccional (unilateral)
   Cuando está seleccionado, la potencia se calcula para una prueba unilateral.
9. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
10. Opcionalmente, puede pulsar **Gráfico** para especificar los valores de "Análisis de potencia de correlación de Pearson para una muestra" en la página 28 (salida de gráfica, valores de gráfico en dos dimensiones y valores de gráfico en tres dimensiones).

    Nota: Gráfico solo está disponible cuando se selecciona **Potencia** como supuesto de prueba.
11. De forma opcional, pulse **Precisión** para calcular el tamaño de muestra en función de los intervalos de confianza especificando los valores de la mitad del intervalo de confianza. Para obtener más información, consulte "Análisis de potencia: Precisión" en la página 38.

Nota: Precisión solo está disponible cuando se selecciona **Tamaño de muestra** como método de **Estimación** de supuesto de prueba y se selecciona el análisis **No direccional (bilateral)** como **Dirección de prueba**.

Análisis de potencia de correlación de Pearson para una muestra

El diálogo **Gráfico** proporciona opciones para controlar los gráficos de salida para ilustrar la potencia bidimensional y tridimensional mediante gráficas. El diálogo también controla los grados de rotación vertical y horizontal para gráficas tridimensionales.

Gráfico de dos dimensiones
Proporciona opciones para controlar la estimación de potencia bidimensional versus los gráficos. De forma predeterminada, los valores están inhabilitados.

Estimación de potencia versus valor de hipótesis nula
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante un gráfico de valores nulos. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus valor de hipótesis alternativa
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante un gráfico de valores alternativos. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus la diferencia entre valores hipotetizados
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de diferencia entre valores hipotetizados. De forma predeterminada, el valor está inhabilitado.

Estimación de potencia versus tamaño de muestra (en pares)
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de tamaño de la muestra. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Rango de gráfico de tamaño de muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor o igual que 4, y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Gráfico tridimensional
Proporciona opciones para controlar la estimación de potencia tridimensional versus gráficos. Este valor está inhabilitado de forma predeterminada.

Estimación de potencia versus tamaño de la muestra
Cuando se selecciona, este valor habilita las opciones siguientes.

en el eje x y la diferencia entre los valores hipotetizados en el eje y
El valor opcional controla la potencia tridimensional mediante el gráfico de tamaño de la muestra (eje x) y la diferencia entre valores hipotetizados(eje y). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

en el eje y y la diferencia entre los valores hipotetizados en el eje x
El valor opcional controla la potencia tridimensional mediante el gráfico de tamaño de la muestra (eje y) y la diferencia entre valores hipotetizados(eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se visualiza el gráfico

Rango de gráfico de tamaño de muestra (en pares)
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia tridimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor o igual que 4, y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia tridimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Estimación de potencia versus valor de hipótesis nula
Cuando se selecciona, este valor habilita las opciones siguientes.

en el eje x y el valor de hipótesis alternativa en el eje y
El valor opcional controla la potencia tridimensional mediante un gráfico de valor alternativo y nulo (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

en el eje y y el valor de hipótesis alternativa en eje x
El valor opcional controla la potencia tridimensional mediante un gráfico de valor nulo y alternativo (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Análisis de potencia de prueba de correlación de Spearman para una muestra

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

El coeficiente de correlación de orden de rangos de Spearman es una estadística no paramétrica basada en rango para medir la relación monotónica entre dos variables que suelen censurarse y no se distribuyen normalmente. La correlación de orden de rango de Spearman es igual a la correlación de Pearson entre los valores de rango de las dos variables, por lo que también varía entre -1 y 1. La detección de la potencia de la prueba de correlación de rango de Spearman es una cuestión importante en el análisis de los datos de series de tiempo hidrológico.

La prueba utiliza el método asintótico de Fisher para estimar la potencia para la correlación de Spearman de una muestra.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Correlaciones** > **Orden de rango de Spearman**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores de potencia de cuadrícula** y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

   Puede obtener información adicional consultando “Análisis de potencia: Valores de cuadrícula” en la página 39.
4. Cuando se selecciona **Potencia** como valor de **Estimación** de supuesto de prueba, especifique el valor correspondiente de **Tamaño de muestra en pares**. El valor debe ser un entero único mayor que 3.
5. Especifique un valor que especifique el valor de hipótesis alternativo del parámetro de correlación en el campo **Parámetro de correlación de Spearman**. El valor debe ser un único numérico entre -1 y 1.

   Nota: Cuando se especifica **Potencia**, el valor de **Parámetro de correlación de Spearman** no puede ser -1 o 1 y no puede ser igual a **Valor nulo**.

6. Opcionalmente, entre un valor que especifique el valor de hipótesis nulo del parámetro de correlación que se probará en el campo **Valor nulo**. El valor debe ser un único numérico entre -1 y 1. El valor predeterminado es 0.

   Nota: Cuando se especifica **Potencia**, **Valor nulo** no puede ser -1 o 1.

7. Opcionalmente, seleccione una opción que determine cómo se estima la varianza asintótica para el análisis de potencia.

   Bonett y Wright
   Estima la varianza sugerida por Bonett y Wright. Ésta es el ajuste predeterminado.

   Fieller, Hartley y Pearson
   Estima la varianza sugerida por Fieller, Hartley y Pearson.

   Caruso y Cliff
   Estima la varianza sugerida por Caruso y Cliff.

8. Seleccione si la prueba es unilateral o bilateral.

   Análisis no direccional (bilateral)
   Cuando está seleccionado, se utiliza una prueba bilateral. Ésta es el ajuste predeterminado.

   Análisis direccional (unilateral)
   Cuando está seleccionado, la potencia se calcula para una prueba unilateral.

9. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
10. Opcionalmente, pulse **Gráfico** para especificar los valores de "Análisis de potencia de correlación de Spearman para una muestra: Gráfico" en la página 31 (salida de gráfica, valores de gráfico en dos dimensiones y valores de gráfico en tres dimensiones).

    Nota: Gráfico solo está disponible cuando se selecciona **Potencia** como supuesto de prueba.

11. De forma opcional, pulse **Precisión** para calcular el tamaño de muestra en función de los intervalos de confianza especificando los valores de la mitad del intervalo de confianza. Para obtener más información, consulte "Análisis de potencia: Precisión" en la página 38.

Nota: Precisión solo está disponible cuando se selecciona **Tamaño de muestra** como método de **Estimación** de supuesto de prueba y se selecciona el análisis **No direccional (bilateral)** como **Dirección de prueba**.

Análisis de potencia de correlación de Spearman para una muestra: Gráfico

El diálogo **Gráfico** proporciona opciones para controlar los gráficos de salida para ilustrar la potencia bidimensional y tridimensional mediante gráficas. El diálogo también controla los grados de rotación vertical y horizontal para gráficas tridimensionales.

Gráfico de dos dimensiones
Proporciona opciones para controlar la estimación de potencia bidimensional versus los gráficos. De forma predeterminada, los valores están inhabilitados.

Estimación de potencia versus valor de hipótesis nula
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante un gráfico de valores nulos. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus valor de hipótesis alternativa
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante un gráfico de valores alternativos. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus la diferencia entre valores hipotetizados
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de diferencia entre valores hipotetizados. De forma predeterminada, el valor está inhabilitado.

Estimación de potencia versus tamaño de muestra (en pares)
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de tamaño de la muestra. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Rango de gráfico de tamaño de muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor o igual que 4, y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Gráfico tridimensional
Proporciona opciones para controlar la estimación de potencia tridimensional versus gráficos. Este valor está inhabilitado de forma predeterminada.

Estimación de potencia versus tamaño de la muestra
Cuando se selecciona, este valor habilita las opciones siguientes.

en el eje x y la diferencia entre los valores hipotetizados en el eje y
El valor opcional controla la potencia tridimensional mediante el gráfico de tamaño de la muestra (eje x) y la diferencia entre valores hipotetizados(eje y). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

en el eje y y la diferencia entre los valores hipotetizados en el eje x
El valor opcional controla la potencia tridimensional mediante el gráfico de tamaño de la muestra (eje y) y la diferencia entre valores hipotetizados(eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se visualiza el gráfico

Rango de gráfico de tamaño de muestra (en pares)
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia tridimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor o igual que 4, y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia tridimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Estimación de potencia versus valor de hipótesis nula
Cuando se selecciona, este valor habilita las opciones siguientes.

en el eje x y el valor de hipótesis alternativa en el eje y
El valor opcional controla la potencia tridimensional mediante un gráfico de valor alternativo y nulo (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

en el eje y y el valor de hipótesis alternativa en eje x
El valor opcional controla la potencia tridimensional mediante un gráfico de valor nulo y alternativo (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Análisis de potencia de prueba de correlación de Pearson parcial

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

La correlación parcial se puede explicar como la asociación entre dos variables aleatorias tras eliminar el efecto de otra variable o de varias de ellas. Es una medida útil si hay confusión. El coeficiente de correlación parcial, similar al Coeficiente de correlación de Pearson, es también una cantidad sin dimensionar que oscila entre -1 y 1.

La prueba utiliza el método asintótico de Fisher para estimar la potencia para la correlación de Pearson de una muestra.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Correlaciones** > **Parcial**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores de potencia de cuadrícula** y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

   Puede obtener información adicional consultando “Análisis de potencia: Valores de cuadrícula” en la página 39.
4. Cuando se selecciona **Potencia** como valor de **Estimación** de supuesto de prueba, especifique el valor correspondiente de **Tamaño de muestra**. El valor debe ser un único entero mayor que 1.
5. Indique un valor que especifique el **Número de variables que se suponen parcializadas**. El valor debe ser un entero único mayor o igual que 0.
6. Indique un valor que especifique el valor de hipótesis alternativo del **Parámetro de correlación parcial**. El valor debe ser un único numérico entre -1 y 1.

   Nota: Cuando se especifica **Potencia**, el valor de **Parámetro de correlación parcial** no puede ser -1 o 1 y no puede ser igual al **Valor nulo**.
7. Opcionalmente, entre un valor que especifica el valor de hipótesis nulo del parámetro de correlación parcial que se probará en el campo **Valor nulo**. El valor debe ser un único numérico entre -1 y 1. El valor predeterminado es 0.

   Nota: Cuando se especifica **Potencia**, **Valor nulo** no puede ser -1 o 1.
8. Seleccione si la prueba es unilateral o bilateral.

Análisis no direccional (bilateral)
Cuando está seleccionado, se utiliza una prueba bilateral. Ésta es el ajuste predeterminado.

Análisis direccional (unilateral)
Cuando está seleccionado, la potencia se calcula para una prueba unilateral.

9. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
10. Opcionalmente, pulse **Gráfico** para especificar los valores de "Análisis de potencia de correlación de Pearson parcial: Gráfico" en la página 34 (salida de gráfica, valores de gráfico en dos dimensiones y valores de gráfico en tres dimensiones).

    Nota: Gráfico solo está disponible cuando se selecciona **Potencia** como supuesto de prueba.
11. De forma opcional, pulse **Precisión** para calcular el tamaño de muestra en función de los intervalos de confianza especificando los valores de la mitad del intervalo de confianza. Para obtener más información, consulte "Análisis de potencia: Precisión" en la página 38.

Nota: Precisión solo está disponible cuando se selecciona **Tamaño de muestra** como método de **Estimación** de supuesto de prueba y se selecciona el análisis **No direccional (bilateral)** como **Dirección de prueba**.

Análisis de potencia de correlación de Pearson parcial: Gráfico

El diálogo **Gráfico** proporciona opciones para controlar los gráficos de salida para ilustrar la potencia bidimensional y tridimensional mediante gráficas. El diálogo también controla los grados de rotación vertical y horizontal para gráficas tridimensionales.

Gráfico de dos dimensiones
Proporciona opciones para controlar la estimación de potencia bidimensional versus los gráficos. De forma predeterminada, los valores están inhabilitados.

Estimación de potencia versus valor de hipótesis nula
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante un gráfico de valores nulos. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus valor de hipótesis alternativa
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante un gráfico de valores alternativos. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus número de variables parcializadas
Cuando está habilitado, este valor opcional controla la potencia bidimensional por número de gráfico de variables de salida parcializadas. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus la diferencia entre valores hipotetizados
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de diferencia entre valores hipotetizados. De forma predeterminada, el valor está inhabilitado.

Estimación de potencia versus tamaño de la muestra
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de tamaño de la muestra. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Rango de gráfico de tamaño de muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor o igual que 4, y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Gráfico tridimensional
Proporciona opciones para controlar la estimación de potencia tridimensional versus gráficos. Este valor está inhabilitado de forma predeterminada.

Estimación de potencia versus tamaño de la muestra
Cuando se selecciona, este valor habilita las opciones siguientes.

en el eje x y la diferencia entre los valores hipotetizados en el eje y
El valor opcional controla la potencia tridimensional mediante el gráfico de tamaño de la muestra (eje x) y la diferencia entre valores hipotetizados(eje y). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

en el eje y y la diferencia entre los valores hipotetizados en el eje x
El valor opcional controla la potencia tridimensional mediante el gráfico de tamaño de la muestra (eje y) y la diferencia entre valores hipotetizados(eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se visualiza el gráfico

Rango de gráfico de tamaño de muestra (en pares)
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia tridimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor o igual que 4, y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia tridimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Estimación de potencia versus valor de hipótesis nula
Cuando se selecciona, este valor habilita las opciones siguientes.

en el eje x y el valor de hipótesis alternativa en el eje y
El valor opcional controla la potencia tridimensional mediante un gráfico de valor alternativo y nulo (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

en el eje y y el valor de hipótesis alternativa en eje x
El valor opcional controla la potencia tridimensional mediante un gráfico de valor nulo y alternativo (eje x). De forma predeterminada, se suprime el gráfico. Cuando se especifica, se muestra el gráfico.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Regresión

Las siguientes características de estadísticas se incluyen en IBM SPSS Estadística Base Edition.

Análisis de potencia de la prueba de regresión lineal univariada

Esta característica requiere IBM SPSS Estadística Base Edition.

El análisis de la potencia desempeña un rol fundamental en el plan, diseño y realización de un plan. Normalmente, el cálculo de la potencia se realiza antes de que se hayan recopilado datos de muestra excepto, posiblemente, en el caso de un estudio piloto pequeño. Una estimación precisa de la potencia puede indicar a los investigadores la probabilidad con la que se detectará una diferencia estadísticamente diferente basándose en un tamaño de muestra finito bajo una hipótesis alternativa real. Si la potencia es demasiado baja, hay pocas probabilidades de detectar una diferencia significativa y son probables resultados no significativos, aunque existan de verdad diferencias reales.

La regresión lineal univariada es un enfoque básico y estándar en el que los investigadores utilizan los valores de varias variables para explicar o predecir valores de un resultado de escala.

La prueba de Análisis de potencia de regresión linea univariada estima la potencia de la prueba *F* del tipo III en modelos de regresión lineal múltiple univariada. Con el tamaño de efecto representado por correlaciones (parciales) múltiples, se proporcionan método para predictores fijos y aleatorios. Para los predictores fijos, la estimación de potencia se basa en la distribución *F* no central. Para los predictores aleatorios, se supone que la variable de destino y los predictores siguen conjuntamente una distribución normal multivariante. En cuyo caso, la estimación de potencia se basa en la distribución del coeficiente de correlación múltiple de la muestra.

1. En los menús seleccione:

   Analizar > **Análisis de potencia** > **Regresión** > **Lineal univariada**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione un valor de **Estimación** de supuesto de prueba (**Tamaño de muestra** o **Potencia**).
3. Cuando se seleccione **Tamaño de muestra**, indique un **Valor de potencia único** como valor de estimación de tamaño de muestra (el valor debe ser un valor único entre 0 y 1) o seleccione **Valores de potencia de cuadrícula** y pulse **Cuadrícula** para ver los tamaños de muestra proyectados para un rango de valores de potencia específicos.

   Puede obtener información adicional consultando “Análisis de potencia: Valores de cuadrícula” en la página 39.
4. Cuando se selecciona **Potencia estimada**, especifique un **Tamaño de muestra** apropiado para el valor de estimación de potencia. El valor debe ser un único entero mayor o igual que el número total de predictores de modelo +2 (cuando **Incluir el término de intersección en el modelo** está habilitado). De lo contrario, el valor debe ser un entero único mayor o igual que el número total de predictores de modelo +1.
5. Especifique el valor del coeficiente de correlación parcial múltiple en el campo **Correlación parcial múltiple de población**. El valor debe ser un valor único entre -1 y 1.

   Nota: Cuando se especifica un valor de **Potencia**, el valor de **Correlación parcial múltiple** no puede ser 0.

   Los valores siguientes están habilitados cuando se selecciona **Correlación parcial múltiple de población**:

   Número total de predictores en el modelo
   Especifique el número de predictores totales o los predictores en el modelo completo (sin incluir la intersección, si procede). El valor debe ser un entero único mayor o igual que 1.

   Número de predictores de prueba
   Especifique el número de predictores de prueba o los predictores en el modelo anidado (sin incluir la intersección, si procede). El valor debe ser mayor o igual que 1, pero no es mayor que el valor **Número total de predictores del modelo**.
6. Especifique **Valores R cuadrado de** coeficientes múltiples de correlación para **Modelo completo** y **Modelo anidado**. Los valores deben ser valores únicos entre 0 y 1.

Nota: Cuando se especifica un valor de **Potencia**, el valor de **Modelo completo** debe ser mayor que el valor **Modelo anidado**.

Los valores siguientes están habilitados cuando se selecciona **Valores R cuadrado de**:

Número total de predictores - Modelo completo
Especifique el número de predictores totales del modelo completo (sin incluir la interceptación, si procede). El valor debe ser un entero único mayor o igual que 1.

Número total de predictores - Modelo anidado
Especifique el número de predictores totales del modelo anidado (sin incluir la interceptación, si procede). El valor debe ser mayor o igual que 1, pero menor que el valor de **Número total de predictores - Modelo completo**.

7. Opcionalmente, especifique el nivel de significación del índice de errores de tipo I en el campo **Nivel de significación**. El valor debe ser un único valor doble entre 0 y 1. El valor predeterminado es 0,05.
8. Opcionalmente, puede seleccionar el valor **Incluir el término de intersección en el modelo**. Este valor está habilitado de forma predeterminada. Cuando no se selecciona, el término de intersección se excluye del análisis de potencia.
9. Puede opcionalmente seleccionar si los predictores de modelo son **Fijos** o **Aleatorios**. **Fijo** es el valor predeterminado.
10. Opcionalmente, puede pulsar **Gráfico** para especificar los valores de "Análisis de potencia de regresión lineal univariada: Gráfico" en la página 37 (salida de gráfica, valores de gráfico en dos dimensiones y valores de gráfico en tres dimensiones).

Nota: Gráfico solo está disponible cuando se selecciona **Potencia** como supuesto de prueba.

Análisis de potencia de regresión lineal univariada: Gráfico

Puede controlar los gráficos que se generan para ilustrar los gráficos de potencia de dos y tres dimensiones. También puede controlar la visualización de las ayudas contextuales y los grados de rotación vertical/horizontal para los gráficos tridimensionales.

Gráfico de dos dimensiones
Proporciona opciones para controlar la estimación de potencia bidimensional versus los gráficos. De forma predeterminada, los valores están inhabilitados.

Estimación de potencia versus la correlación parcial múltiple
Cuando habilita esta opción, este valor opcional controla la potencia bidimensional mediante un gráfico de coeficiente de correlación parcial. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Estimación de potencia versus tamaño de la muestra
Cuando está habilitado, este valor opcional controla la potencia bidimensional mediante el gráfico de tamaño de la muestra. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Rango de gráfico de tamaño de muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor o igual que 4, y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Gráfico tridimensional
Proporciona opciones para controlar la estimación de potencia tridimensional versus los gráficos, los valores de rotación vertical/horizontal y el rango de gráfico de tamaño de muestra especificado por el usuario. Este valor está inhabilitado de forma predeterminada.

Estimación de potencia versus tamaño de la muestra
Cuando está habilitado, este valor opcional controla la potencia tridimensional mediante gráficos de tamaño de la muestra. De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

en el eje x, y la correlación parcial múltiple en el eje y
Este valor opcional controla la potencia tridimensional mediante el tamaño de la muestra (eje x) y el gráfico de coeficiente de correlación parcial múltiple (eje y). De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

en el eje y, y la correlación parcial múltiple en el eje x
Este valor opcional controla la potencia tridimensional mediante el tamaño de la muestra (eje x) y el gráfico de coeficiente de correlación parcial múltiple (eje y). De forma predeterminada, el valor está inhabilitado. Cuando se selecciona, este valor muestra el gráfico.

Rango de gráfico de tamaño de muestra
Cuando está seleccionado, las opciones de límite inferior y superior están disponibles. Cuando no se especifican ningún valor de entero para los campos **Límite inferior** o **Límite superior**, se utiliza el rango de gráfico predeterminado.

Límite inferior
Controla el límite inferior para la estimación de potencia bidimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor o igual que 4, y no puede ser mayor que el valor del **Límite superior**.

Límite superior
Controla el límite superior para la estimación de potencia bidimensional mediante el gráfico de tamaño de la muestra. El valor debe ser mayor que el valor de **Límite inferior** y no puede ser mayor que 5000.

Rotación vertical
El valor opcional establece los grafos de rotación vertical (en el sentido de las agujas del reloj desde la izquierda) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico verticalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 10.

Rotación horizontal
El valor opcional establece los grados de rotación horizontal (en el sentido de las agujas del reloj desde la parte frontal) para el gráfico tridimensional. Puede utilizar el ratón para rotar el gráfico horizontalmente. El valor entra en vigor cuando se solicita el gráfico tridimensional. El valor debe ser un solo valor de entero menor o igual que 359. El valor predeterminado es 325.

Análisis de potencia: Precisión

El diálogo **Precisión** está disponible para cada procedimiento de análisis de potencia (excepto para la regresión lineal univariada) y proporciona opciones para estimar el tamaño de la muestra en función de la mitad del intervalo de confianza especificado. El diálogo está disponible cuando se especifican un **Valor de potencia único** o **Valores de potencia de cuadrícula** para la estimación de **Tamaño de muestra** y cuando el valor **Dirección de prueba** se establece en **No direccional (análisis bilateral)**.

Nota: El valor **Dirección de prueba** no se aplica a ANOVA de un factor.

cuando se seleccionan las opciones **Calcular el tamaño de muestra** y **Valores de potencia de cuadrícula** (pulse el control **Cuadrícula** para visualizar el diálogo).

Tipo(s) de intervalo de confianza
Seleccione los tipos de intervalos de confianza adecuados y si se aplica la corrección de continuidad.

Nota: Tipo(s) de intervalo de confianza sólo están disponibles para los procedimientos de proporción de análisis de potencia (Proporción de una muestra, Prueba binomial de muestras relacionadas y Prueba binomial de muestras independientes).

Especifique la mitad del ancho del intervalo de confianza
Estima el tamaño de muestra basándose en el valor de medio ancho del intervalo de confianza. Especifique un valor en el rango de 0 a 1. Para la prueba binomial de una muestra, el valor debe estar en el rango de 0 a 0,5.

Nota: Se ignoran los valores duplicados.

- Pulse **Añadir** para añadir el valor de media anchura especificado a la lista.
- Resalte un valor de media anchura existente y pulse **Cambiar** para actualizar el valor.
- Resalte un valor de media anchura existente y pulse **Eliminar** para eliminar el valor de la lista.

Análisis de potencia: Valores de cuadrícula

El diálogo **Valores de cuadrícula** proporciona opciones para especificar un rango de valores de **POWER** para fines de visualización de los tamaños de muestra proyectados en un formato de cuadrícula para cada valor de rango de **POWER** especificado.

El diálogo **Valores de cuadrícula** está disponible en cada procedimiento de análisis de potencia cuando se seleccionan las opciones **Tamaño de muestra estimado** y **Valores de potencia de cuadrícula** (pulse el control **Cuadrícula** para mostrar el diálogo).

Especificar una sola potencia
Cuando está seleccionado, se requiere al menos un valor para ejecutar el análisis. Están permitidos varios valores y cada valor debe pertenecer a `[0, 1]`. Puede especificar varios valores separando cada valor con un espacio en blanco (o espacios en banco). Utilice los controles **Añadir**, **Cambiar** y **Eliminar** para trabajar con los valores de la lista de valores de potencia.

Todos los valores deben ser exclusivos (los valores duplicados no están permitidos).

Especificar rango de potencia
Cuando se selecciona, se puede especificar un rango de valores de potencia desde un valor **Inicial** (`value1`) hasta un valor **Final** (`value2`) con el incremento de **Por** (`value3`). Si se especifica, sólo se permite un conjunto válido de [`value1 TO value2 BY value3`]. Debe satisfacer ese `0 ≤ value1 ≤ value2 ≤ 1`. En los casos en que `value1 = value2`, equivale a especificar un único `value1`, independientemente de `value3`.

Nota: Las opciones **Especificar potencia única** y **Especificar rango de potencia** son independientes; puede seleccionar una opción individual o ambas opciones.

Metaanálisis

El metaanálisis es el análisis de los datos obtenidos de una recopilación de estudios que responden a preguntas de investigación similares. Estos estudios se denominan estudios primarios. El metaanálisis utiliza métodos estadísticos para producir una estimación global de un efecto, explorar la heterogeneidad entre estudios e investigar el impacto del sesgo de publicación o, de forma más general, los efectos de los estudios pequeños en los resultados finales.

IBM SPSS Estadística da soporte a tamaños de efecto estándar y tamaños de efecto genéricos (precalculados) tanto para datos binarios (como la razón de probabilidad logarítmica) como para datos continuos (como por ejemplo g de Hedges). Durante el paso de la declaración de meta-análisis se especifica la información de meta-análisis (como los tamaños de efecto específicos del estudio y sus correspondientes errores estándar, y el modelo y el método de meta-análisis). La información se utiliza automáticamente en todos los metaanálisis posteriores.

Se admiten modelos de meta-análisis de efectos fijos, de efectos aleatorios y de efecto común. Según el modelo de meta-análisis elegido, están disponibles los distintos métodos (por ejemplo, la varianza inversa y Mantel-Haenszel) para los modelos de efecto común y de efectos fijos. También hay varios

estimadores distintos disponibles para el parámetro de varianza entre estudios para el modelo de efectos aleatorios.

Meta-análisis continuo

El procedimiento Meta-análisis continuo realiza el meta-análisis con resultados continuos en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Ejemplo

En el pasado, se llevaron a cabo varios estudios de investigación para investigar un medicamento novedoso pero controvertido para tratar la diabetes de tipo II. Se afirmaba que el medicamento oral reducía el nivel de glucosa en sangre después de las comidas. Se recopilaron datos de diferentes fuentes de investigación de 1979 a 1986.

Una de las directoras de la investigación pretendía extraer la inferencia estadística sobre el efecto del medicamento oral. Debido al hecho de que se generaron datos a partir de diferentes estudios, propuso la idea de sintetizar los resultados de todos los estudios para conocer el efecto global e identificar las fuentes subyacentes de variación en los resultados.

Estadísticos

Nivel de confianza, método iterativo, subdivisión por pasos, tolerancia de convergencia, medias muestrales, varianza muestral, desviación estándar, tamaño de efecto estimado, *d* de Cohen, *g* de Hedges, Delta de Glass, diferencia de medias, análisis acumulado, método de estimación, recortar y rellenar, prueba basada en regresión, modelo de efectos aleatorios, modelo de efectos fijos, estimador de la máxima verosimilitud restringida, estimador de Bayesiano empírico, estimador de Hedges, estimador de Hunter-Schmidt, estimador de DerSimonian-Laird, estimador de Sidik-Jonkman, ajuste de error estándar de Knapp-Hartung, ajuste de error estándar de Knapp-Hartung, coeficientes, prueba basada en regresión EGGER, intersección, modelo multiplicativo, parámetro de dispersión multiplicativa, estimador cuadrático, prueba de homogeneidad, medidas de heterogeneidad, intervalo de predicción, error estándar estimado, valor *p* estimado, tamaño de efecto global acumulado, ponderación de estudio estimado.

Obtención de un análisis de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Datos en bruto...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. En la sección **Grupo de tratamiento**, seleccione una variable **Tamaño de estudio** para representar el tamaño de la muestra para el grupo de tratamiento. La variable seleccionada debe ser numérica (no se admiten variables de cadena).
3. Seleccione una variable **Media** para representar las medias muestrales para el grupo de tratamiento. La variable seleccionada debe ser numérica (no se admiten variables de cadena).
4. Seleccione **Desviación estándar** para especificar la desviación estándar de la muestra o **Varianza** para especificar la varianza muestral y, a continuación, seleccione una variable para representar la desviación estándar/varianza para el grupo de tratamiento.
5. En la sección **Grupo de control**, seleccione una variable **Tamaño de estudio** para representar el tamaño de la muestra para el grupo de control. La variable seleccionada debe ser numérica (no se admiten variables de cadena).
6. Seleccione una variable **Media** para representar las medias muestrales para el grupo de control. La variable seleccionada debe ser numérica (no se admiten variables de cadena).
7. Seleccione **Desviación estándar** para especificar la desviación estándar de la muestra o **Varianza** para especificar la varianza muestral y, a continuación, seleccione una variable para representar la desviación estándar/varianza para el grupo de control.
8. Si lo desea, seleccione las variables **ID de estudio** y/o **Etiqueta de estudio**. La variable **ID de estudio** seleccionada no puede ser la misma que la variable **Etiqueta de estudio** seleccionada.

9. Si lo desea, seleccione un valor de **Tamaño de efecto**.

   d de Cohen
   El valor predeterminado estima la *d* de Cohen. Cuando se selecciona **Error estándar ajustado**, el ajuste estima la *d* de Cohen y su varianza utilizando una fórmula alternativa dividida por $2(N_{treatment} + N_{control} - 2)$.

   g de Hedges
   Estima la *g* de Hedges. Cuando se selecciona **Error estándar ajustado**, el ajuste estima la *g* de Hedges y su varianza utilizando una fórmula alternativa dividida por $2(N_{treatment} + N_{control} - 3.94)$.

   Delta de Glass
   Estima el Delta de Glass en función del grupo de control. Cuando se selecciona **Estandarizado basado en el grupo de tratamiento**, el Delta de Glass se estandariza en función de la desviación estándar del grupo de tratamiento.

   Diferencia de medias no estandarizada
   Estima la diferencia de medias presuponiendo que las dos desviaciones estándar de población son iguales. Cuando se selecciona **Varianzas de grupo desiguales**, la diferencia de medias se estima presuponiendo que las dos desviaciones estándar de población no son iguales.

10. Si lo desea, seleccione un valor de **Modelo**. Cuando los valores de **Recortar y rellenar** están habilitados, el valor también controla el modelo que se utiliza en la agrupación en el análisis de recorte y relleno. Cuando los valores de **Sesgo** están habilitados, el valor también controla el modelo que utiliza la prueba basada en regresión.

    Efectos aleatorios
    El valor predeterminado crea el modelo de efectos aleatorios.

    Efectos fijos
    Crea el modelo de efectos fijos.

11. Si lo desea, puede:

    - Haz clic en **Criterios...** para especificar los criterios generales.
    - Pulse **Análisis** para especificar el subgrupo y el análisis acumulado.
    - Pulse **Inferencia** para especificar los métodos de estimación.
    - Pulse **Contraste** para controlar la prueba de contraste.
    - Pulse **Sesgo** para acceder al sesgo de publicación realizando la prueba basada en regresión EGGER.
    - Pulse **Recortar y rellenar** para implementar el análisis de recorte y relleno de sesgo de publicación.
    - Pulse **Imprimir** para controlar los resultados de la tabla.
    - Pulse **Guardar** para guardar las estadísticas estimadas en el conjunto de datos activo.
    - Pulse **Gráfico** para especificar qué gráficos incluir en la salida.

12. Pulse **Aceptar**.

Meta-análisis continuo: Criterios

El diálogo **Criterios** proporciona valores para especificar criterios para el metaanálisis con resultados continuos sobre los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Intervalo de confianza
El valor opcional especifica el nivel de confianza. El valor debe estar comprendido por un valor numérico entre 0 y 100. El valor predeterminado es 95.

Falta el ámbito de datos
Los valores opcionales controlan cómo gestiona el procedimiento la falta de datos.

Excluir casos según análisis
El valor predeterminado incluye todos los casos con datos suficientes en las variables utilizadas en cada análisis concreto.

Excluir casos según lista
Incluye todos los casos con datos suficientes en todas las variables en todos los análisis especificados por el procedimiento.

Valores perdidos de usuario
Los valores opcionales controlan cómo se tratan los valores perdidos de usuario.

Excluir
El valor predeterminado trata los valores perdidos de usuario como válidos.

Incluir
No tiene en cuenta las designaciones de valor perdido y trata los valores perdidos de usuario como válidos.

Iteraciones

Número máximo de iteraciones
El valor opcional especifica el número máximo de iteraciones en los métodos iterativos. El valor está disponible cuando se utilizan métodos iterativos. El valor debe ser un entero positivo único. El valor predeterminado es 100. El valor 0 indica que no se realizan iteraciones.

Máximo de subdivisión por pasos
El valor opcional especifica la máxima subdivisión por pasos en los métodos iterativos. El valor está disponible cuando se utilizan métodos iterativos. El valor debe ser un entero positivo único. El valor predeterminado es 5. Un valor de 0 significa que no se aplica la subdivisión por pasos.

Convergencia
El valor opcional especifica la tolerancia de convergencia. El valor debe ser un valor positivo único. El valor está disponible cuando se utilizan métodos iterativos. El valor predeterminado es 1E-6.

Definición de criterios de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis continuo**, pulse **Criterios**.
3. Seleccione y defina los valores de criterios adecuados.
4. Pulse **Continuar**.

Meta-análisis continuo: Análisis

El diálogo **Análisis** proporciona valores para especificar el análisis acumulado y de subgrupos para el meta-análisis con resultados continuos en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Análisis de subgrupos
Seleccione una variable que evoque el análisis de subgrupos. La variable no puede ser la misma que la variable definida para **Análisis acumulado**.

Análisis acumulado
Seleccione una variable que evoque el análisis acumulado y sobre la que se realice el meta-análisis acumulado. La variable no puede ser la misma que la variable definida para **Análisis de subgrupos**. Cuando se selecciona **Ascendente**, el análisis acumulado se basa en la variable especificada en orden ascendente. Cuando se selecciona **Descendente**, el análisis acumulado se basa en la variable especificada en orden descendente.

Estadísticos acumulados
Proporciona opciones para guardar el tamaño de efecto global acumulado estimado. Los valores solo están disponibles cuando se selecciona una variable de **Análisis acumulado**.

Tamaño de efecto acumulado
Guarda el tamaño de efecto global acumulado estimado.

Error estándar
Guarda el error estándar estimado del tamaño de efecto global acumulado.

Límite inferior de intervalo de confianza
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto global acumulado.

Límite superior de intervalo de confianza
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto global acumulado.

Valor P
Guarda el valor *p* estimado del tamaño de efecto global acumulado.

Destino
Proporciona opciones para especificar un archivo de datos o un conjunto de datos de guardado. Cuando se selecciona **Conjunto de datos**, puede especificar un nuevo nombre de conjunto de datos (puede conservar el nombre de conjunto de datos predeterminado). Una vez seleccionado **el archivo de datos**, haga clic en **Examinar...** para seleccionar un nombre de archivo y una ubicación.

Definición de valores de análisis de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis continuo**, pulse **Análisis**.
3. Seleccione y defina los valores de análisis adecuados.
4. Pulse **Continuar**.

Meta-análisis continuo: Inferencia

El diálogo **Inferencia** proporciona valores para especificar los métodos de estimación para el metaanálisis con resultados continuos sobre los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Nota: El diálogo **Inferencia** solo está disponible cuando se selecciona un modelo de **Efectos aleatorios**.

Estimador
Proporciona los valores para especificar el estimador.

Notas:

- Cuando se especifican valores de Recorte y relleno, también controla el estimador que se utiliza mediante la agrupación en el análisis de recorte y relleno.
- Cuando se especifican valores de sesgo, también controla el estimador que utiliza la prueba basada en regresión.

Máxima verosimilitud restringida (REML)
El valor predeterminado aplica el método iterativo y calcula el estimador de máxima verosimilitud restringida.

Máxima verosimilitud (ML)
Aplica el método iterativo y calcula el estimador de máxima verosimilitud.

Bayes empírico
Aplica el método iterativo y calcula el estimador de Bayes empírico.

Hedges
Aplica el método no iterativo y calcula el estimador de Hedges.

Hunter-Schmidt
Aplica el método no iterativo y calcula el estimador de Hunter-Schmidt.

DerSimonian-Laird
Aplica el método no iterativo y calcula el estimador de DerSimonian-Laird.

Sidik-Jonkman
Aplica el método no iterativo y calcula el estimador de Sidik-Jonkman.

Ajuste del error estándar
Proporciona los valores que controlan si se aplica el ajuste de error estándar de Knapp-Hartung.

Notas:

- Cuando se especifican valores de Recorte y relleno, también controla el ajuste del error estándar que se utiliza mediante la agrupación en el análisis de recorte y relleno.
- Cuando se especifican valores de sesgo, también controla el ajuste del error estándar que utiliza la prueba basada en regresión.

Sin ajuste
El valor predeterminado no aplica el ajuste.

Aplicar el ajuste de Knapp-Hartung
Aplica el método de ajuste de Knapp-Hartung.

Aplicar el ajuste de Knapp-Hartung truncado
Aplica el método de ajuste de Knapp-Hartung y trunca el valor si es menor que 1 al estimar la matriz de varianza-covarianza.

Definición de valores de inferencia de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis continuo**, pulse **Inferencia**.
3. Seleccione y defina los valores de inferencia adecuados.
4. Pulse **Continuar**.

Meta-análisis continuo: Contraste

El diálogo **Contraste** proporciona valores para controlar la prueba de contraste para el meta-análisis con resultados continuos en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Variables
La lista muestra todas las variables de conjunto de datos disponibles. Seleccione variables de la lista y muévalas a la lista **Contraste(s)**.

Contraste(s)
La lista identifica los coeficientes que se almacenan como variables en el conjunto de datos activo. Se permiten varias variables. Las variables de cadena no están soportadas.

Valores de coeficiente de entrada de usuario
Proporciona valores para especificar los coeficientes de un contraste suministrado por el usuario. Solo se permiten valores numéricos. Para formular una prueba de contraste válida, el número de valores especificados debe coincidir con el de los estudios válidos.

Definición de valores de contraste de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis continuo**, pulse **Contraste**.
3. Seleccione y defina los valores de contraste adecuados.
4. Pulse **Continuar**.

Meta-análisis continuo: Sesgo

El diálogo **Sesgo** proporciona valores para habilitar el sesgo de publicación realizando la prueba de regresión Eggers para el meta-análisis con resultados continuos en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Prueba basada en regresión Egger
La selección de este valor habilita el sesgo de publicación realizando la prueba basada en regresión Eggers.

Variables
La lista proporciona todas las variables de conjunto de datos disponibles.

Covariables
Las variables seleccionadas de la lista **Variables** se tratan como covariables. Se permiten varias covariables.

Factores
Las variables seleccionadas de la lista **Variables** se tratan como factores. Se permiten varios factores.

Incluir la intersección en la regresión
Controla el término de intersección en la prueba basada en regresión.

Incluir parámetro de dispersión en el modelo de efectos fijos
Controla el ajuste de modelo multiplicativo e introduce el parámetro de dispersión multiplicativo en el análisis. El valor solo está disponible cuando se selecciona un modelo de efectos fijos.

Estimar estadísticas en función de la distribución t
Controla la distribución utilizada en las pruebas basadas en regresión. El valor está habilitado de forma predeterminada, lo que estima las estadísticas basadas en la distribución *t*. Cuando el valor no está seleccionado, las estadísticas se estiman en función de la distribución normal.

Definición de valores de sesgo de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis continuo**, pulse **Sesgo**.
3. Seleccione y defina los valores de sesgo adecuados.
4. Pulse **Continuar**.

Meta-análisis continuo: Recortar y rellenar

El diálogo **Recorte y relleno** proporciona valores para implementar el análisis de recorte y relleno del sesgo de publicación para el metaanálisis con resultados continuos sobre los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Número estimado de estudios perdidos
Controla el análisis de recorte y relleno de sesgo de publicación. La selección de este valor habilita los demás valores del diálogo.

Lado para imputar estudios
Proporciona opciones para especificar el lado del gráfico de embudo en el que se imputan los estudios perdidos.

Determinado por las inclinaciones de prueba Egger
El valor predeterminado determina el lado en función de la inclinación estimada de la prueba de Egger.

Izquierda
Imputa el lado izquierdo del gráfico de embudo.

Derecha
Imputa el lado derecho del gráfico de embudo.

Método
Especifica el método para estimar el número de estudios perdidos.

Lineal
El valor predeterminado calcula el estimador lineal.

Ejecute
Calcula el estimador de ejecución.

Cuadrático
Calcula el estimador cuadrático.

Proceso de iteración
Proporciona valores para especificar el estimador de iteración y el ajuste del error estándar.

Modelo de efectos fijos
Cuando se selecciona, se utiliza un modelo de efectos fijos y las opciones de estimación de iteración y ajuste de error estándar no están disponibles.

Modelo de efectos aleatorios
Cuando se selecciona, se utiliza un modelo de efectos aleatorios y están disponibles los valores siguientes.

Estimador
Proporciona valores para especificar el estimador de iteración

Máxima verosimilitud restringida (REML)
El valor predeterminado aplica el método iterativo y calcula el estimador de máxima verosimilitud restringida.

Máxima verosimilitud (ML)
Aplica el método iterativo y calcula el estimador de máxima verosimilitud.

Bayes empírico
Aplica el método iterativo y calcula el estimador de Bayes empírico.

Hedges
Aplica el método no iterativo y calcula el estimador de Hedges.

Hunter-Schmidt
Aplica el método no iterativo y calcula el estimador de Hunter-Schmidt.

DerSimonian-Laird
Aplica el método no iterativo y calcula el estimador de DerSimonian-Laird.

Sidik-Jonkman
Aplica el método no iterativo y calcula el estimador de Sidik-Jonkman.

Ajuste del error estándar
Proporciona valores para controlar si se aplica o no el ajuste de error estándar de Knapp-Hartung a las iteraciones del algoritmo de recorte y relleno.

Sin ajuste
El valor predeterminado no aplica el ajuste.

Aplicar el ajuste de Knapp-Hartung
Aplica el método de ajuste de Knapp-Hartung.

Aplicar el ajuste de Knapp-Hartung truncado
Aplica el método de ajuste de Knapp-Hartung y trunca el valor si es menor que 1 al estimar la matriz de varianza-covarianza.

Definición de valores de recorte y relleno de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis continuo**, pulse **Recortar y rellenar**.
3. Seleccione y defina los valores adecuados de recorte y relleno.

4. Pulse **Continuar**.

Meta-análisis continuo: Imprimir

El diálogo **Imprimir** proporciona valores para controlar los resultados de tabla para el metaanálisis con resultados continuos sobre los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Homogeneidad/Heterogeneidad
Proporciona valores para controlar la homogeneidad y la heterogeneidad.

Prueba de homogeneidad
Cuando se selecciona se proporciona la prueba de homogeneidad correspondiente en los resultados.

Medidas de heterogeneidad
Cuando se selecciona, se proporcionan las medidas de heterogeneidad en los resultados.

Tamaños de efecto
Proporciona los valores de tamaño de efecto siguientes.

Estudios individuales
Controla la visualización de los estudios individuales. Cuando se selecciona, se proporcionan los resultados correspondientes.

Tamaños de efecto acumulados
Controla la visualización del análisis acumulado. Cuando se selecciona, se proporcionan los resultados correspondientes. El valor solo está disponible cuando se selecciona una variable de **Análisis acumulado** en el diálogo **Análisis**.

Intervalo de predicción en el modelo de efectos aleatorios
Controla la visualización del intervalo de predicción. Cuando se selecciona, se proporcionan los resultados correspondientes. El valor solo está disponible cuando se especifica un modelo de efectos aleatorios.

Definición de valores de impresión de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis continuo**, pulse **Imprimir**.
3. Seleccione y defina los valores de impresión adecuados.
4. Pulse **Continuar**.

Meta-análisis continuo: Guardar

El diálogo **Guardar** proporciona valores para guardar las estadísticas estimadas en el conjunto de datos activo para el meta-análisis con resultados continuos en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Estudios individuales
Proporciona opciones para guardar el tamaño de efecto estimado.

Tamaño de efecto individual
Guarda el tamaño de efecto estimado.

Error estándar
Guarda el error estándar estimado del tamaño de efecto.

Límite inferior de intervalo de confianza
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto.

Límite superior de intervalo de confianza
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto.

Valor P
Guarda el valor *p* estimado del tamaño de efecto.

Ponderación de estudio
Guarda la ponderación de estudio estimada.

Porcentaje de ponderación de estudios
Guarda la ponderación de estudio normalizada como un porcentaje.

Definición de valores de guardado de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis continuo**, pulse **Guardar**.
3. Seleccione y defina los valores de guardado adecuados de estadísticas estimadas.
4. Pulse **Continuar**.

Meta-análisis continuo: Gráfico

El diálogo **Gráfico** proporciona valores para los siguientes tipos de gráfico:

- "Gráfico de bosque" en la página 78
- "Gráfico de bosque acumulado" en la página 79
- "Gráfico de burbujas" en la página 80
- "Gráfico de embudo" en la página 81
- "Gráfico de Galbraith" en la página 82

Tamaño de efecto de meta-análisis continuo

El procedimiento Metaanálisis Continuo Tamaño de efecto realiza el metaanálisis con resultados continuos cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Ejemplo
En el pasado, se llevaron a cabo varios estudios de investigación para investigar un medicamento novedoso pero controvertido para tratar la diabetes de tipo II. Se afirmaba que el medicamento oral reducía el nivel de glucosa en sangre después de las comidas. Se recopilaron datos de diferentes fuentes de investigación de 1979 a 1986.

Una de las directoras de la investigación pretendía extraer la inferencia estadística sobre el efecto del medicamento oral. Debido al hecho de que se generaron datos a partir de diferentes estudios, propuso la idea de sintetizar los resultados de todos los estudios para conocer el efecto global e identificar las fuentes subyacentes de variación en los resultados.

Estadísticos
Nivel de confianza, método iterativo, subdivisión por pasos, tolerancia de convergencia, medias muestrales, varianza muestral, desviación estándar, tamaño de efecto estimado, *d* de Cohen, *g* de Hedges, Delta de Glass, diferencia de medias, análisis acumulado, método de estimación, recortar y rellenar, prueba basada en regresión, modelo de efectos aleatorios, modelo de efectos fijos, estimador de la máxima verosimilitud restringida, estimador de Bayesiano empírico, estimador de Hedges, estimador de Hunter-Schmidt, estimador de DerSimonian-Laird, estimador de Sidik-Jonkman, ajuste de error estándar de Knapp-Hartung, ajuste de error estándar de Knapp-Hartung, coeficientes, prueba basada en regresión EGGER, intersección, modelo multiplicativo, parámetro de dispersión multiplicativa, estimador cuadrático, prueba de homogeneidad, medidas de heterogeneidad, intervalo de predicción, error estándar estimado, valor *p* estimado, tamaño de efecto global acumulado, ponderación de estudio estimado,

Obtención de un análisis de tamaño de efecto de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Tamaño de efecto calculado previamente**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una variable **Tamaño de efecto** que indique el tamaño de efecto. La variable seleccionada debe ser numérica (no se admiten variables de cadena).
3. Seleccione una **Desviación estándar** para especificar la desviación estándar del tamaño de efecto o **Varianza** para especificar la varianza del tamaño de efecto y, a continuación, seleccione una variable para representar la desviación estándar o la varianza del tamaño de efecto.
4. Si lo desea, seleccione las variables **ID de estudio** y/o **Etiqueta de estudio**. La variable **ID de estudio** seleccionada no puede ser la misma que la variable **Etiqueta de estudio** seleccionada.
5. Si lo desea, seleccione un valor de **Modelo**. Cuando los valores de **Recortar y rellenar** están habilitados, el valor también controla el modelo que se utiliza en la agrupación en el análisis de recorte y relleno. Cuando los valores de **Sesgo** están habilitados, el valor también controla el modelo que utiliza la prueba basada en regresión.

   Efectos aleatorios
   El valor predeterminado crea el modelo de efectos aleatorios.

   Efectos fijos
   Crea el modelo de efectos fijos.
6. Si lo desea, puede:
   - Haz clic en **Criterios...** para especificar los criterios generales.
   - Pulse **Análisis** para especificar el subgrupo y el análisis acumulado.
   - Pulse **Inferencia** para especificar los métodos de estimación.
   - Pulse **Contraste** para controlar la prueba de contraste.
   - Pulse **Sesgo** para acceder al sesgo de publicación realizando la prueba basada en regresión EGGER.
   - Pulse **Recortar y rellenar** para implementar el análisis de recorte y relleno de sesgo de publicación.
   - Pulse **Imprimir** para controlar los resultados de la tabla.
   - Pulse **Guardar** para guardar las estadísticas estimadas en el conjunto de datos activo.
   - Pulse **Gráfico** para especificar qué gráficos incluir en la salida.
7. Pulse **Aceptar**.

Tamaño de efecto de meta-análisis continuo: Criterios

El diálogo **Criterios** proporciona valores para especificar los criterios para el metaanálisis con resultados continuos cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Intervalo de confianza
El valor opcional especifica el nivel de confianza. El valor debe estar comprendido por un valor numérico entre 0 y 100. El valor predeterminado es 95.

Falta el ámbito de datos
Los valores opcionales controlan cómo gestiona el procedimiento la falta de datos.

Excluir casos según análisis
El valor predeterminado incluye todos los casos con datos suficientes en las variables utilizadas en cada análisis concreto.

Excluir casos según lista
Incluye todos los casos con datos suficientes en todas las variables en todos los análisis especificados por el procedimiento.

Valores perdidos de usuario
Los valores opcionales controlan cómo se tratan los valores perdidos de usuario.

Excluir
El valor predeterminado trata los valores perdidos de usuario como válidos.

Incluir
No tiene en cuenta las designaciones de valor perdido y trata los valores perdidos de usuario como válidos.

Iteraciones

Número máximo de iteraciones
El valor opcional especifica el número máximo de iteraciones en los métodos iterativos. El valor está disponible cuando se utilizan métodos iterativos. El valor debe ser un entero positivo único. El valor predeterminado es 100. El valor 0 indica que no se realizan iteraciones.

Máximo de subdivisión por pasos
El valor opcional especifica la máxima subdivisión por pasos en los métodos iterativos. El valor está disponible cuando se utilizan métodos iterativos. El valor debe ser un entero positivo único. El valor predeterminado es 5. Un valor de 0 significa que no se aplica la subdivisión por pasos.

Convergencia
El valor opcional especifica la tolerancia de convergencia. El valor debe ser un valor positivo único. El valor está disponible cuando se utilizan métodos iterativos. El valor predeterminado es 1E-6.

Definición de criterios de tamaño de efecto de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Tamaño de efecto calculado previamente**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis continuo**, pulse **Criterios**.
3. Seleccione y defina los valores de criterios adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis continuo: Análisis

El diálogo **Análisis** proporciona valores para especificar el análisis de subgrupo y acumulativo para el metaanálisis con resultados continuos cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Análisis de subgrupos
Seleccione una variable que evoque el análisis de subgrupos. La variable no puede ser la misma que la variable definida para **Análisis acumulado**.

Análisis acumulado
Seleccione una variable que evoque el análisis acumulado y sobre la que se realice el meta-análisis acumulado. La variable no puede ser la misma que la variable definida para **Análisis de subgrupos**. Cuando se selecciona **Ascendente**, el análisis acumulado se basa en la variable especificada en orden ascendente. Cuando se selecciona **Descendente**, el análisis acumulado se basa en la variable especificada en orden descendente.

Estadísticos acumulados
Proporciona opciones para guardar el tamaño de efecto global acumulado estimado. Los valores solo están disponibles cuando se selecciona una variable de **Análisis acumulado**.

Tamaño de efecto acumulado
Guarda el tamaño de efecto global acumulado estimado.

Error estándar
Guarda el error estándar estimado del tamaño de efecto global acumulado.

Límite inferior de intervalo de confianza
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto global acumulado.

Límite superior de intervalo de confianza
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto global acumulado.

Valor P
Guarda el valor *p* estimado del tamaño de efecto global acumulado.

Destino
Proporciona opciones para especificar un archivo de datos o un conjunto de datos de guardado. Cuando se selecciona **Conjunto de datos**, puede especificar un nuevo nombre de conjunto de datos (puede conservar el nombre de conjunto de datos predeterminado). Una vez seleccionado **el archivo de datos**, haga clic en **Examinar...** para seleccionar un nombre de archivo y una ubicación.

Definición de valores de análisis de tamaño de efecto de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Tamaño de efecto calculado previamente**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis continuo**, pulse **Análisis**.
3. Seleccione y defina los valores de análisis adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis continuo: Inferencia

El diálogo **Inferencia** proporciona valores para especificar los métodos de estimación para el metaanálisis con resultados continuos cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Nota: El diálogo **Inferencia** solo está disponible cuando se selecciona un modelo de **Efectos aleatorios**.

Estimador
Proporciona los valores para especificar el estimador.

Notas:

- Cuando se especifican valores de Recorte y relleno, también controla el estimador que se utiliza mediante la agrupación en el análisis de recorte y relleno.
- Cuando se especifican valores de sesgo, también controla el estimador que utiliza la prueba basada en regresión.

Máxima verosimilitud restringida (REML)
El valor predeterminado aplica el método iterativo y calcula el estimador de máxima verosimilitud restringida.

Máxima verosimilitud (ML)
Aplica el método iterativo y calcula el estimador de máxima verosimilitud.

Bayes empírico
Aplica el método iterativo y calcula el estimador de Bayes empírico.

Hedges
Aplica el método no iterativo y calcula el estimador de Hedges.

Hunter-Schmidt
Aplica el método no iterativo y calcula el estimador de Hunter-Schmidt.

DerSimonian-Laird
Aplica el método no iterativo y calcula el estimador de DerSimonian-Laird.

Sidik-Jonkman
Aplica el método no iterativo y calcula el estimador de Sidik-Jonkman.

Ajuste del error estándar
Proporciona los valores que controlan si se aplica el ajuste de error estándar de Knapp-Hartung.

Notas:

- Cuando se especifican valores de Recorte y relleno, también controla el ajuste del error estándar que se utiliza mediante la agrupación en el análisis de recorte y relleno.
- Cuando se especifican valores de sesgo, también controla el ajuste del error estándar que utiliza la prueba basada en regresión.

Sin ajuste
El valor predeterminado no aplica el ajuste.

Aplicar el ajuste de Knapp-Hartung
Aplica el método de ajuste de Knapp-Hartung.

Aplicar el ajuste de Knapp-Hartung truncado
Aplica el método de ajuste de Knapp-Hartung y trunca el valor si es menor que 1 al estimar la matriz de varianza-covarianza.

Definición de valores de inferencia de tamaño de efecto de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Tamaño de efecto calculado previamente**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis continuo**, pulse **Inferencia**.
3. Seleccione y defina los valores de inferencia adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis continuo: Contraste

El diálogo **Contraste** proporciona valores para controlar la prueba de contraste para el metaanálisis con resultados continuos cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Variables
La lista muestra todas las variables de conjunto de datos disponibles. Seleccione variables de la lista y muévalas a la lista **Contraste(s)**.

Contraste(s)
La lista identifica los coeficientes que se almacenan como variables en el conjunto de datos activo. Se permiten varias variables. Las variables de cadena no están soportadas.

Valores de coeficiente de entrada de usuario
Proporciona valores para especificar los coeficientes de un contraste suministrado por el usuario. Solo se permiten valores numéricos. Para formular una prueba de contraste válida, el número de valores especificados debe coincidir con el de los estudios válidos.

Definición de valores de contraste de tamaño de efecto de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Tamaño de efecto calculado previamente**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis continuo**, pulse **Contraste**.
3. Seleccione y defina los valores de contraste adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis continuo: Sesgo

El cuadro de diálogo **Sesgo** proporciona valores para habilitar el sesgo de publicación realizando la prueba de regresión de Eggers para el meta-análisis con resultados continuos cuando los datos de tamaño de efecto calculados previamente se proporcionan en el conjunto de datos activo.

Prueba basada en regresión Egger
La selección de este valor habilita el sesgo de publicación realizando la prueba basada en regresión Eggers.

Variables
La lista proporciona todas las variables de conjunto de datos disponibles.

Covariables
Las variables seleccionadas de la lista **Variables** se tratan como covariables. Se permiten varias covariables.

Factores
Las variables seleccionadas de la lista **Variables** se tratan como factores. Se permiten varios factores.

Incluir la intersección en la regresión
Controla el término de intersección en la prueba basada en regresión.

Incluir parámetro de dispersión en el modelo de efectos fijos
Controla el ajuste de modelo multiplicativo e introduce el parámetro de dispersión multiplicativo en el análisis. El valor solo está disponible cuando se selecciona un modelo de efectos fijos.

Estimar estadísticas en función de la distribución t
Controla la distribución utilizada en las pruebas basadas en regresión. El valor está habilitado de forma predeterminada, lo que estima las estadísticas basadas en la distribución *t*. Cuando el valor no está seleccionado, las estadísticas se estiman en función de la distribución normal.

Definición de valores de sesgo de tamaño de efecto de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Tamaño de efecto calculado previamente**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis continuo**, pulse **Sesgo**.
3. Seleccione y defina los valores de sesgo adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis continuo: Recortar y rellenar

El diálogo **Recorte y relleno** proporciona valores para implementar el análisis de recorte y relleno del sesgo de publicación para el metaanálisis con resultados continuos cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Número estimado de estudios perdidos
Controla el análisis de recorte y relleno de sesgo de publicación. La selección de este valor habilita los demás valores del diálogo.

Lado para imputar estudios
Proporciona opciones para especificar el lado del gráfico de embudo en el que se imputan los estudios perdidos.

- **Determinado por las inclinaciones de prueba Egger**
  El valor predeterminado determina el lado en función de la inclinación estimada de la prueba de Egger.
- **Izquierda**
  Imputa el lado izquierdo del gráfico de embudo.
- **Derecha**
  Imputa el lado derecho del gráfico de embudo.

Método
Especifica el método para estimar el número de estudios perdidos.

- **Lineal**
  El valor predeterminado calcula el estimador lineal.
- **Ejecute**
  Calcula el estimador de ejecución.
- **Cuadrático**
  Calcula el estimador cuadrático.

Proceso de iteración
Proporciona valores para especificar el estimador de iteración y el ajuste del error estándar.

Modelo de efectos fijos
Cuando se selecciona, se utiliza un modelo de efectos fijos y las opciones de estimación de iteración y ajuste de error estándar no están disponibles.

Modelo de efectos aleatorios
Cuando se selecciona, se utiliza un modelo de efectos aleatorios y están disponibles los valores siguientes.

Estimador
Proporciona valores para especificar el estimador de iteración

Máxima verosimilitud restringida (REML)
El valor predeterminado aplica el método iterativo y calcula el estimador de máxima verosimilitud restringida.

Máxima verosimilitud (ML)
Aplica el método iterativo y calcula el estimador de máxima verosimilitud.

Bayes empírico
Aplica el método iterativo y calcula el estimador de Bayes empírico.

Hedges
Aplica el método no iterativo y calcula el estimador de Hedges.

Hunter-Schmidt
Aplica el método no iterativo y calcula el estimador de Hunter-Schmidt.

DerSimonian-Laird
Aplica el método no iterativo y calcula el estimador de DerSimonian-Laird.

Sidik-Jonkman
Aplica el método no iterativo y calcula el estimador de Sidik-Jonkman.

Ajuste del error estándar
Proporciona valores para controlar si se aplica o no el ajuste de error estándar de Knapp-Hartung a las iteraciones del algoritmo de recorte y relleno.

Sin ajuste
El valor predeterminado no aplica el ajuste.

Aplicar el ajuste de Knapp-Hartung
Aplica el método de ajuste de Knapp-Hartung.

Aplicar el ajuste de Knapp-Hartung truncado
Aplica el método de ajuste de Knapp-Hartung y trunca el valor si es menor que 1 al estimar la matriz de varianza-covarianza.

Definición de valores de recorte y relleno de tamaño de efecto de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Tamaño de efecto calculado previamente**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis continuo**, pulse **Recortar y rellenar**.
3. Seleccione y defina los valores adecuados de recorte y relleno.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis continuo: Imprimir

El diálogo **Imprimir** proporciona valores para controlar los resultados de tabla para el metaanálisis con resultados continuos cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Homogeneidad/Heterogeneidad
Proporciona valores para controlar la homogeneidad y la heterogeneidad.

Prueba de homogeneidad
Cuando se selecciona se proporciona la prueba de homogeneidad correspondiente en los resultados.

Medidas de heterogeneidad
Cuando se selecciona, se proporcionan las medidas de heterogeneidad en los resultados.

Tamaños de efecto
Proporciona los valores de tamaño de efecto siguientes.

Estudios individuales
Controla la visualización de los estudios individuales. Cuando se selecciona, se proporcionan los resultados correspondientes.

Tamaños de efecto acumulados
Controla la visualización del análisis acumulado. Cuando se selecciona, se proporcionan los resultados correspondientes. El valor solo está disponible cuando se selecciona una variable de **Análisis acumulado** en el diálogo **Análisis**.

Intervalo de predicción en el modelo de efectos aleatorios
Controla la visualización del intervalo de predicción. Cuando se selecciona, se proporcionan los resultados correspondientes. El valor solo está disponible cuando se especifica un modelo de efectos aleatorios.

Definición de valores de impresión de tamaño de efecto de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Tamaño de efecto calculado previamente**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis continuo**, pulse **Imprimir**.
3. Seleccione y defina los valores de impresión adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis continuo: Guardar

El cuadro de diálogo **Guardar** proporciona valores para guardar las estadísticas estimadas en el conjunto de datos activo para el meta-análisis con resultados continuos cuando los datos de tamaño de efecto calculados previamente se proporcionan en el conjunto de datos activo.

Estudios individuales
Proporciona opciones para guardar el tamaño de efecto estimado.

Error estándar
Guarda el error estándar estimado del tamaño de efecto. El valor solo está disponible cuando no se especifica una variable de **Tamaño de efecto**.

Límite inferior de intervalo de confianza
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto.

Límite superior de intervalo de confianza
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto.

Valor P
Guarda el valor *p* estimado del tamaño de efecto.

Ponderación de estudio
Guarda la ponderación de estudio estimada.

Porcentaje de ponderación de estudios
Guarda la ponderación de estudio normalizada como un porcentaje.

Definición de valores de guardado de tamaño de efecto de meta-análisis continuo

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados continuos** > **Tamaño de efecto calculado previamente**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis continuo**, pulse **Guardar**.
3. Seleccione y defina los valores de guardado adecuados de estadísticas estimadas.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis continuo: Gráfico

El diálogo **Gráfico** proporciona valores para los siguientes tipos de gráfico:

- "Gráfico de bosque" en la página 78
- "Gráfico de bosque acumulado" en la página 79
- "Gráfico de burbujas" en la página 80
- "Gráfico de embudo" en la página 81
- "Gráfico de Galbraith" en la página 82

Meta-análisis binario

El procedimiento Meta-análisis binario realiza el meta-análisis con resultados binarios en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Ejemplo

En el pasado, se llevaron a cabo varios estudios de investigación para investigar un medicamento novedoso pero controvertido para tratar la diabetes de tipo II. Se afirmaba que el medicamento oral reducía el nivel de glucosa en sangre después de las comidas. Se recopilaron datos de diferentes fuentes de investigación de 1979 a 1986.

Una de las directoras de la investigación pretendía extraer la inferencia estadística sobre el efecto del medicamento oral. Debido al hecho de que se generaron datos a partir de diferentes estudios, propuso la idea de sintetizar los resultados de todos los estudios para conocer el efecto global e identificar las fuentes subyacentes de variación en los resultados.

Estadísticos

Intervalo de confianza, Razón de probabilidad de registro, Razón de probabilidad de registro Peto, Razón de riesgo de registro, diferencia de riesgo, efectos aleatorios, efectos fijos, varianza inversa, Mantel-Haenszel, iteraciones, subdivisión por pasos, convergencia, estadísticas acumuladas, tamaño de efecto acumulado, máxima verosimilitud restringida, MVR, máxima verosimilitud, MV, Bayesiano empírico, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, Prueba Egger, Prueba Harbord, Prueba Peters, intersección en la regresión, parámetro de dispersión, homogeneidad, heterogeneidad, estadísticas exponenciadas, error estándar, valor *p*, ponderación de estudio.

Obtención de un análisis de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados binarios** > **Datos en bruto...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. En la sección **Grupo de tratamiento**, seleccione una variable **Éxito** para representar los recuentos de "éxito" para el grupo de tratamiento. La variable seleccionada debe ser numérica (no se admiten variables de cadena).
3. Seleccione una variable **Error** para representar los recuentos de "error" para el grupo de tratamiento. La variable seleccionada debe ser numérica (no se admiten variables de cadena).

4. En la sección **Grupo de control**, seleccione una variable **Éxito** para representar los recuentos de "éxito" para el grupo de control. La variable seleccionada debe ser numérica (no se admiten variables de cadena).
5. Seleccione una variable **Error** para representar los recuentos de "error" para el grupo de control. La variable seleccionada debe ser numérica (no se admiten variables de cadena).
6. Si lo desea, seleccione las variables **ID de estudio** y/o **Etiqueta de estudio**. La variable **ID de estudio** seleccionada no puede ser la misma que la variable **Etiqueta de estudio** seleccionada.
7. Si lo desea, seleccione un valor de **Tamaño de efecto**. Las opciones disponibles son **Razón de probabilidad de registro**, **Razón de probabilidad de registro Peto**, **Razón de riesgo de registro** y **Diferencia de riesgo**.
8. Si lo desea, seleccione un valor de **Modelo**. Cuando los valores de **Recortar y rellenar** están habilitados, el valor también controla el modelo que se utiliza en la agrupación en el análisis de recorte y relleno. Cuando los valores de **Sesgo** están habilitados, el valor también controla el modelo que utiliza la prueba basada en regresión.

   Efectos aleatorios
   El valor predeterminado crea el modelo de efectos aleatorios.

   Efectos fijos
   Crea el modelo de efectos fijos. **Varianza inversa** estima la ponderación de varianza inversa. **Mantel-Haenszel** estima la ponderación de Mantel-Haenszel.
9. Si lo desea, puede:
   - Haz clic en **Criterios...** para especificar los criterios generales.
   - Pulse **Análisis** para especificar el subgrupo y el análisis acumulado.
   - Pulse **Inferencia** para especificar los métodos de estimación.
   - Pulse **Contraste** para controlar la prueba de contraste.
   - Pulse **Sesgo** para acceder al sesgo de publicación realizando la prueba basada en regresión EGGER.
   - Pulse **Recortar y rellenar** para implementar el análisis de recorte y relleno de sesgo de publicación.
   - Pulse **Imprimir** para controlar los resultados de la tabla.
   - Pulse **Guardar** para guardar las estadísticas estimadas en el conjunto de datos activo.
   - Pulse **Gráfico** para especificar qué gráficos incluir en la salida.
10. Pulse **Aceptar**.

Meta-análisis binario: Criterios

El diálogo **Criterios** proporciona valores para especificar los criterios para el meta-análisis con resultados binarios en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Intervalo de confianza
El valor opcional especifica el nivel de confianza. El valor debe estar comprendido por un valor numérico entre 0 y 100. El valor predeterminado es 95.

Falta el ámbito de datos
Los valores opcionales controlan cómo gestiona el procedimiento la falta de datos.

Excluir casos según análisis
El valor predeterminado incluye todos los casos con datos suficientes en las variables utilizadas en cada análisis concreto.

Excluir casos según lista
Incluye todos los casos con datos suficientes en todas las variables en todos los análisis especificados por el procedimiento.

Valores perdidos de usuario
Los valores opcionales controlan cómo se tratan los valores perdidos de usuario.

Excluir
El valor predeterminado trata los valores perdidos de usuario como válidos.

Incluir
No tiene en cuenta las designaciones de valor perdido y trata los valores perdidos de usuario como válidos.

Iteraciones

Número máximo de iteraciones
El valor opcional especifica el número máximo de iteraciones en los métodos iterativos. El valor está disponible cuando se utilizan métodos iterativos. El valor debe ser un entero positivo único. El valor predeterminado es 100. El valor 0 indica que no se realizan iteraciones.

Máximo de subdivisión por pasos
El valor opcional especifica la máxima subdivisión por pasos en los métodos iterativos. El valor está disponible cuando se utilizan métodos iterativos. El valor debe ser un entero positivo único. El valor predeterminado es 5. Un valor de 0 significa que no se aplica la subdivisión por pasos.

Convergencia
El valor opcional especifica la tolerancia de convergencia. El valor debe ser un valor positivo único. El valor está disponible cuando se utilizan métodos iterativos. El valor predeterminado es 1E-6.

Ajustar recuentos cero
Los valores opcionales controlan cómo ajustar los datos de recuento cero, si procede, en la estimación de tamaño de efecto.

Ajustar cada valor de solo aquellos estudios que contengan al menos un cero
El valor predeterminado ajusta solo los datos de recuento cero.

Ajustar cada valor de todos los estudios solo si existe al menos un estudio que contiene un valor cero
Ajusta todas las frecuencias cuando existe al menos un recuento cero en un estudio.

Ajustar cada valor de todos los estudios
Ajusta todos los datos independientemente de la existencia de recuentos cero.

No hacer ajustes y mantener cero(s)
No realiza ningún ajuste de datos.

Valor añadido
El valor opcional especifica el valor añadido a los datos de recuento cero. El valor predeterminado es 0.5. El valor especificado debe ser un único valor numérico mayor que 0 y menor o igual que 1.

Definición de criterios de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados binarios** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis binario**, pulse **Criterios**.
3. Seleccione y defina los valores de criterios adecuados.
4. Pulse **Continuar**.

Meta-análisis binario: Análisis

El diálogo **Análisis** proporciona valores para especificar el análisis acumulado y de subgrupos para el meta-análisis con resultados binarios en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Análisis de subgrupos
Seleccione una variable que evoque el análisis de subgrupos. La variable no puede ser la misma que la variable definida para **Análisis acumulado**.

Análisis acumulado
Seleccione una variable que evoque el análisis acumulado y sobre la que se realice el meta-análisis acumulado. La variable no puede ser la misma que la variable definida para **Análisis de subgrupos**. Cuando se selecciona **Ascendente**, el análisis acumulado se basa en la variable especificada en orden ascendente. Cuando se selecciona **Descendente**, el análisis acumulado se basa en la variable especificada en orden descendente.

Estadísticos acumulados
Proporciona opciones para guardar el tamaño de efecto global acumulado estimado. Los valores solo están disponibles cuando se selecciona una variable de **Análisis acumulado**.

Tamaño de efecto acumulado
Guarda el tamaño de efecto global acumulado estimado.

Tamaño de efecto acumulado (formato exponenciado)
Guarda el tamaño de efecto acumulado estimado en formato exponencial.

Error estándar
Guarda el error estándar estimado del tamaño de efecto global acumulado.

Límite inferior de intervalo de confianza
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto global acumulado.

Límite superior de intervalo de confianza
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto global acumulado.

Límite inferior de intervalo de confianza (formato exponenciado)
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto global acumulado en formato exponencial.

Límite superior de intervalo de confianza (formato exponenciado)
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto global acumulado en formato exponencial.

Valor P
Guarda el valor *p* estimado del tamaño de efecto global acumulado.

Destino
Proporciona opciones para especificar un archivo de datos o un conjunto de datos de guardado. Cuando se selecciona **Conjunto de datos**, puede especificar un nuevo nombre de conjunto de datos (puede conservar el nombre de conjunto de datos predeterminado). Una vez seleccionado **el archivo de datos**, haga clic en **Examinar...** para seleccionar un nombre de archivo y una ubicación.

Definición de valores de análisis de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados binarios** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis binario**, pulse **Análisis**.
3. Seleccione y defina los valores de análisis adecuados.
4. Pulse **Continuar**.

Meta-análisis binario: Inferencia

El diálogo **Inferencia** proporciona valores para especificar los métodos de estimación para el metaanálisis con resultados binarios sobre los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Nota: El diálogo **Inferencia** solo está disponible cuando se selecciona un modelo de **Efectos aleatorios**.

Estimador
Proporciona los valores para especificar el estimador.

Notas:

- Cuando se especifican valores de Recorte y relleno, también controla el estimador que se utiliza mediante la agrupación en el análisis de recorte y relleno.
- Cuando se especifican valores de sesgo, también controla el estimador que utiliza la prueba basada en regresión.

Máxima verosimilitud restringida (REML)
El valor predeterminado aplica el método iterativo y calcula el estimador de máxima verosimilitud restringida.

Máxima verosimilitud (ML)
Aplica el método iterativo y calcula el estimador de máxima verosimilitud.

Bayes empírico
Aplica el método iterativo y calcula el estimador de Bayes empírico.

Hedges
Aplica el método no iterativo y calcula el estimador de Hedges.

Hunter-Schmidt
Aplica el método no iterativo y calcula el estimador de Hunter-Schmidt.

DerSimonian-Laird
Aplica el método no iterativo y calcula el estimador de DerSimonian-Laird.

Sidik-Jonkman
Aplica el método no iterativo y calcula el estimador de Sidik-Jonkman.

Ajuste del error estándar
Proporciona los valores que controlan si se aplica el ajuste de error estándar de Knapp-Hartung.

Notas:

- Cuando se especifican valores de Recorte y relleno, también controla el ajuste del error estándar que se utiliza mediante la agrupación en el análisis de recorte y relleno.
- Cuando se especifican valores de sesgo, también controla el ajuste del error estándar que utiliza la prueba basada en regresión.

Sin ajuste
El valor predeterminado no aplica el ajuste.

Aplicar el ajuste de Knapp-Hartung
Aplica el método de ajuste de Knapp-Hartung.

Aplicar el ajuste de Knapp-Hartung truncado
Aplica el método de ajuste de Knapp-Hartung y trunca el valor si es menor que 1 al estimar la matriz de varianza-covarianza.

Definición de valores de inferencia de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados binarios** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis binario**, pulse **Inferencia**.
3. Seleccione y defina los valores de inferencia adecuados.
4. Pulse **Continuar**.

Meta-análisis binario: Contraste

El diálogo **Contraste** proporciona valores para controlar la prueba de contraste para el meta-análisis con resultados binarios en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Variables
La lista muestra todas las variables de conjunto de datos disponibles. Seleccione variables de la lista y muévalas a la lista **Contraste(s)**.

Contraste(s)
La lista identifica los coeficientes que se almacenan como variables en el conjunto de datos activo. Se permiten varias variables. Las variables de cadena no están soportadas.

Valores de coeficiente de entrada de usuario
Proporciona valores para especificar los coeficientes de un contraste suministrado por el usuario. Solo se permiten valores numéricos. Para formular una prueba de contraste válida, el número de valores especificados debe coincidir con el de los estudios válidos.

Visualizar estadísticas exponenciadas
Controla la inclusión de los estadísticos en formato exponencial. Cuando se selecciona el valor, los resultados incluyen los estadísticos en formato exponencial, incluido el tamaño de efecto exponencial y los límites de intervalo de confianza. El valor está disponible cuando el **Tamaño de efecto** se especifica como **Razón de probabilidad de registro**, **Razón de probabilidad de registro Peto** o **Razón de riesgo de registro**.

Definición de valores de contraste de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados binarios** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis binario**, pulse **Contraste**.
3. Seleccione y defina los valores de contraste adecuados.
4. Pulse **Continuar**.

Meta-análisis binario: Sesgo

El diálogo **Sesgo** proporciona valores para habilitar el sesgo de publicación realizando pruebas basadas en regresión para el meta-análisis con resultados binarios en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Pruebas basadas en regresión
Proporciona opciones para especificar las pruebas basadas en regresión. Se puede seleccionar más de una prueba.

Prueba de Egger
Cuando está seleccionado, realiza la prueba Egger.

Prueba de Harbord
Cuando está seleccionado, realiza la prueba Harbord. La prueba está disponible cuando el **Tamaño de efecto** se especifica como **Razón de probabilidad de registro** o **Razón de riesgo de registro**.

Prueba de Peters
Cuando está seleccionado, realiza la prueba Peters. La prueba está disponible cuando el **Tamaño de efecto** se especifica como **Razón de probabilidad de registro**.

Variables
La lista proporciona todas las variables de conjunto de datos disponibles.

Covariables
Las variables seleccionadas de la lista **Variables** se tratan como covariables. Se permiten varias covariables.

Factores
Las variables seleccionadas de la lista **Variables** se tratan como factores. Se permiten varios factores.

Incluir la intersección en la regresión
Controla el término de intersección en la prueba basada en regresión.

Incluir parámetro de dispersión en el modelo de efectos fijos
Controla el ajuste de modelo multiplicativo e introduce el parámetro de dispersión multiplicativo en el análisis. El valor solo está disponible cuando se selecciona un modelo de efectos fijos.

Estimar estadísticas en función de la distribución t
Controla la distribución utilizada en las pruebas basadas en regresión. El valor está habilitado de forma predeterminada, lo que estima las estadísticas basadas en la distribución *t*. Cuando el valor no está seleccionado, las estadísticas se estiman en función de la distribución normal.

Definición de valores de sesgo de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados binarios** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis binario**, pulse **Sesgo**.
3. Seleccione y defina los valores de sesgo adecuados.
4. Pulse **Continuar**.

Meta-análisis binario: Recortar y rellenar

El diálogo **Recorte y relleno** proporciona valores para implementar el análisis de recorte y relleno del sesgo de publicación para el metaanálisis con resultados binarios sobre los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Número estimado de estudios perdidos
Controla el análisis de recorte y relleno de sesgo de publicación. La selección de este valor habilita los demás valores del diálogo.

Lado para imputar estudios
Proporciona opciones para especificar el lado del gráfico de embudo en el que se imputan los estudios perdidos.

Determinado por las inclinaciones de prueba Egger
El valor predeterminado determina el lado en función de la inclinación estimada de la prueba de Egger.

Izquierda
Imputa el lado izquierdo del gráfico de embudo.

Derecha
Imputa el lado derecho del gráfico de embudo.

Método
Especifica el método para estimar el número de estudios perdidos.

Lineal
El valor predeterminado calcula el estimador lineal.

Ejecute
Calcula el estimador de ejecución.

Cuadrático
Calcula el estimador cuadrático.

Proceso de iteración
Proporciona valores para especificar el estimador de iteración y el ajuste del error estándar.

Modelo de efectos fijos
Cuando se selecciona, se utiliza un modelo de efectos fijos y las opciones de estimación de iteración y ajuste de error estándar no están disponibles.

Modelo de efectos aleatorios
Cuando se selecciona, se utiliza un modelo de efectos aleatorios y están disponibles los valores siguientes.

Estimador
Proporciona valores para especificar el estimador de iteración

Máxima verosimilitud restringida (REML)
El valor predeterminado aplica el método iterativo y calcula el estimador de máxima verosimilitud restringida.

Máxima verosimilitud (ML)
Aplica el método iterativo y calcula el estimador de máxima verosimilitud.

Bayes empírico
Aplica el método iterativo y calcula el estimador de Bayes empírico.

Hedges
Aplica el método no iterativo y calcula el estimador de Hedges.

Hunter-Schmidt
Aplica el método no iterativo y calcula el estimador de Hunter-Schmidt.

DerSimonian-Laird
Aplica el método no iterativo y calcula el estimador de DerSimonian-Laird.

Sidik-Jonkman
Aplica el método no iterativo y calcula el estimador de Sidik-Jonkman.

Ajuste del error estándar
Proporciona valores para controlar si se aplica o no el ajuste de error estándar de Knapp-Hartung a las iteraciones del algoritmo de recorte y relleno.

Sin ajuste
El valor predeterminado no aplica el ajuste.

Aplicar el ajuste de Knapp-Hartung
Aplica el método de ajuste de Knapp-Hartung.

Aplicar el ajuste de Knapp-Hartung truncado
Aplica el método de ajuste de Knapp-Hartung y trunca el valor si es menor que 1 al estimar la matriz de varianza-covarianza.

Definición de valores de recorte y relleno de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados binarios** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis binario**, pulse **Recortar y rellenar**.
3. Seleccione y defina los valores adecuados de recorte y relleno.
4. Pulse **Continuar**.

Meta-análisis binario: Imprimir

El diálogo **Imprimir** proporciona valores para controlar los resultados de tabla para el metaanálisis con resultados binarios sobre los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Homogeneidad/Heterogeneidad
Proporciona valores para controlar la homogeneidad y la heterogeneidad.

Prueba de homogeneidad
Cuando se selecciona se proporciona la prueba de homogeneidad correspondiente en los resultados.

Medidas de heterogeneidad
Cuando se selecciona, se proporcionan las medidas de heterogeneidad en los resultados.

Tamaños de efecto
Proporciona los valores de tamaño de efecto siguientes.

Estudios individuales
Controla la visualización de los estudios individuales. Cuando se selecciona, se proporcionan los resultados correspondientes.

Tamaños de efecto acumulados
Controla la visualización del análisis acumulado. Cuando se selecciona, se proporcionan los resultados correspondientes. El valor solo está disponible cuando se selecciona una variable de **Análisis acumulado** en el diálogo **Análisis**.

Intervalo de predicción en el modelo de efectos aleatorios
Controla la visualización del intervalo de predicción. Cuando se selecciona, se proporcionan los resultados correspondientes. El valor solo está disponible cuando se especifica un modelo de efectos aleatorios.

Visualizar estadísticas exponenciadas
Controla la inclusión de los estadísticos en formato exponencial. Cuando se selecciona el valor, se generan los estadísticos en formato exponencial, incluido el tamaño de efecto exponencial y los límites de intervalo de confianza. El proceso se aplica a las tablas de predicción y estimación de tamaño de efecto opcionales y necesarias. El valor está disponible cuando el **Tamaño de efecto** se especifica como **Razón de probabilidad de registro**, **Razón de probabilidad de registro Peto** o **Razón de riesgo de registro**.

Definición de valores de impresión de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados binarios** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis binario**, pulse **Imprimir**.
3. Seleccione y defina los valores de impresión adecuados.
4. Pulse **Continuar**.

Meta-análisis binario: Guardar

El diálogo **Guardar** proporciona valores para guardar las estadísticas estimadas en el conjunto de datos activo para el meta-análisis con resultados binarios en los datos en bruto que se proporcionan en el conjunto de datos activo para la estimación del tamaño de efecto.

Estudios individuales
Proporciona opciones para guardar el tamaño de efecto estimado.

Tamaño de efecto individual
Guarda el tamaño de efecto estimado.

Tamaño de efecto individual (formato exponenciado)
Guarda el tamaño de efecto estimado en formato exponencial.

Error estándar
Guarda el error estándar estimado del tamaño de efecto.

Límite inferior de intervalo de confianza
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto.

Límite superior de intervalo de confianza
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto.

Límite inferior de intervalo de confianza (formato exponenciado)
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto en formato exponencial.

Límite superior de intervalo de confianza (formato exponenciado)
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto en formato exponencial.

Valor P
Guarda el valor *p* estimado del tamaño de efecto.

Ponderación de estudio
Guarda la ponderación de estudio estimada.

Porcentaje de ponderación de estudios
Guarda la ponderación de estudio normalizada como un porcentaje.

Definición de valores de guardado de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados binarios** > **Datos en bruto...**
2. En el cuadro de diálogo de **Meta-análisis binario**, pulse **Guardar**.
3. Seleccione y defina los valores de guardado adecuados de estadísticas estimadas.
4. Pulse **Continuar**.

Meta-análisis binario: Gráfico

El diálogo **Gráfico** proporciona valores para los siguientes tipos de gráfico:

- “Gráfico de bosque” en la página 78
- “Gráfico de bosque acumulado” en la página 79
- “Gráfico de burbujas” en la página 80
- “Gráfico de embudo” en la página 81
- “Gráfico de Galbraith” en la página 82
- “Gráfico de L'Abb ´e” en la página 83

Tamaño de efecto de meta-análisis binario

El procedimiento Tamaño de efecto binario de meta-análisis realiza meta-análisis con resultados binarios cuando los datos de tamaño de efecto calculados previamente se proporcionan en el conjunto de datos activo.

Ejemplo
En el pasado, se llevaron a cabo varios estudios de investigación para investigar un medicamento novedoso pero controvertido para tratar la diabetes de tipo II. Se afirmaba que el medicamento oral reducía el nivel de glucosa en sangre después de las comidas. Se recopilaron datos de diferentes fuentes de investigación de 1979 a 1986.

Una de las directoras de la investigación pretendía extraer la inferencia estadística sobre el efecto del medicamento oral. Debido al hecho de que se generaron datos a partir de diferentes estudios, propuso la idea de sintetizar los resultados de todos los estudios para conocer el efecto global e identificar las fuentes subyacentes de variación en los resultados.

Estadísticos
Intervalo de confianza, Razón de probabilidad de registro, Razón de probabilidad de registro Peto, Razón de riesgo de registro, diferencia de riesgo, efectos aleatorios, efectos fijos, varianza inversa, Mantel-Haenszel, iteraciones, subdivisión por pasos, convergencia, estadísticas acumuladas, tamaño de efecto acumulado, máxima verosimilitud restringida, MVR, máxima verosimilitud, MV, Bayesiano empírico, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, Prueba Egger, Prueba Harbord, Prueba Peters, intersección en la regresión, parámetro de dispersión, homogeneidad, heterogeneidad, estadísticas exponenciadas, error estándar, valor *p*, ponderación de estudio.

Obtención de un análisis de tamaño de efecto de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados binarios** > **Tamaño de efecto calculado previamente...**

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione una variable **Tamaño de efecto** que indique el tamaño de efecto. La variable seleccionada debe ser numérica (no se admiten variables de cadena).
3. Seleccione una **Desviación estándar** para especificar la desviación estándar del tamaño de efecto o **Varianza** para especificar la varianza del tamaño de efecto y, a continuación, seleccione una variable para representar la desviación estándar o la varianza del tamaño de efecto.
4. Si lo desea, seleccione las variables **ID de estudio** y/o **Etiqueta de estudio**. La variable **ID de estudio** seleccionada no puede ser la misma que la variable **Etiqueta de estudio** seleccionada.
5. Si lo desea, seleccione un valor de **Tamaño de efecto**. Las opciones disponibles son **Razón de probabilidad de registro**, **Razón de probabilidad de registro Peto**, **Razón de riesgo de registro** y **Diferencia de riesgo**.
6. Si lo desea, seleccione un valor de **Modelo**. Cuando los valores de **Recortar y rellenar** están habilitados, el valor también controla el modelo que se utiliza en la agrupación en el análisis de recorte y relleno. Cuando los valores de **Sesgo** están habilitados, el valor también controla el modelo que utiliza la prueba basada en regresión.

   Efectos aleatorios
   El valor predeterminado crea el modelo de efectos aleatorios.

   Efectos fijos
   Crea el modelo de efectos fijos.
7. Si lo desea, puede:
   - Haz clic en **Criterios...** para especificar los criterios generales.
   - Pulse **Análisis** para especificar el subgrupo y el análisis acumulado.
   - Pulse **Inferencia** para especificar los métodos de estimación.
   - Pulse **Contraste** para controlar la prueba de contraste.
   - Pulse **Sesgo** para acceder al sesgo de publicación realizando la prueba basada en regresión EGGER.
   - Pulse **Recortar y rellenar** para implementar el análisis de recorte y relleno de sesgo de publicación.
   - Pulse **Imprimir** para controlar los resultados de la tabla.
   - Pulse **Guardar** para guardar las estadísticas estimadas en el conjunto de datos activo.
   - Pulse **Gráfico** para especificar qué gráficos incluir en la salida.
8. Pulse **Aceptar**.

Tamaño de efecto de meta-análisis binario: Criterios

El diálogo **Criterios** proporciona valores para especificar los criterios para el metaanálisis con resultados binarios cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Intervalo de confianza
El valor opcional especifica el nivel de confianza. El valor debe estar comprendido por un valor numérico entre 0 y 100. El valor predeterminado es 95.

Falta el ámbito de datos
Los valores opcionales controlan cómo gestiona el procedimiento la falta de datos.

Excluir casos según análisis
El valor predeterminado incluye todos los casos con datos suficientes en las variables utilizadas en cada análisis concreto.

Excluir casos según lista
Incluye todos los casos con datos suficientes en todas las variables en todos los análisis especificados por el procedimiento.

Valores perdidos de usuario
Los valores opcionales controlan cómo se tratan los valores perdidos de usuario.

Excluir
El valor predeterminado trata los valores perdidos de usuario como válidos.

Incluir
No tiene en cuenta las designaciones de valor perdido y trata los valores perdidos de usuario como válidos.

Iteraciones

Número máximo de iteraciones
El valor opcional especifica el número máximo de iteraciones en los métodos iterativos. El valor está disponible cuando se utilizan métodos iterativos. El valor debe ser un entero positivo único. El valor predeterminado es 100. El valor 0 indica que no se realizan iteraciones.

Máximo de subdivisión por pasos
El valor opcional especifica la máxima subdivisión por pasos en los métodos iterativos. El valor está disponible cuando se utilizan métodos iterativos. El valor debe ser un entero positivo único. El valor predeterminado es 5. Un valor de 0 significa que no se aplica la subdivisión por pasos.

Convergencia
El valor opcional especifica la tolerancia de convergencia. El valor debe ser un valor positivo único. El valor está disponible cuando se utilizan métodos iterativos. El valor predeterminado es 1E-6.

Definición de criterios de tamaño de efecto de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados binarios** > **Tamaño de efecto calculado previamente...**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis binario**, pulse **Criterios**.
3. Seleccione y defina los valores de criterios adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis binario: Análisis

El cuadro de diálogo **Análisis** proporciona valores para especificar el análisis acumulado y de subgrupos para el meta-análisis con resultados binarios cuando los datos de tamaño de efecto calculados previamente se proporcionan en el conjunto de datos activo.

Análisis de subgrupos
Seleccione una variable que evoque el análisis de subgrupos. La variable no puede ser la misma que la variable definida para **Análisis acumulado**.

Análisis acumulado
Seleccione una variable que evoque el análisis acumulado y sobre la que se realice el meta-análisis acumulado. La variable no puede ser la misma que la variable definida para **Análisis de subgrupos**. Cuando se selecciona **Ascendente**, el análisis acumulado se basa en la variable especificada en orden ascendente. Cuando se selecciona **Descendente**, el análisis acumulado se basa en la variable especificada en orden descendente.

Estadísticos acumulados
Proporciona opciones para guardar el tamaño de efecto global acumulado estimado. Los valores solo están disponibles cuando se selecciona una variable de **Análisis acumulado**.

Tamaño de efecto acumulado
Guarda el tamaño de efecto global acumulado estimado.

Tamaño de efecto acumulado (formato exponenciado)
Guarda el tamaño de efecto acumulado estimado en formato exponencial.

Error estándar
Guarda el error estándar estimado del tamaño de efecto global acumulado.

Límite inferior de intervalo de confianza
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto global acumulado.

Límite superior de intervalo de confianza
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto global acumulado.

Límite inferior de intervalo de confianza (formato exponenciado)
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto global acumulado en formato exponencial.

Límite superior de intervalo de confianza (formato exponenciado)
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto global acumulado en formato exponencial.

Valor P
Guarda el valor *p* estimado del tamaño de efecto global acumulado.

Destino
Proporciona opciones para especificar un archivo de datos o un conjunto de datos de guardado. Cuando se selecciona **Conjunto de datos**, puede especificar un nuevo nombre de conjunto de datos (puede conservar el nombre de conjunto de datos predeterminado). Una vez seleccionado **el archivo de datos**, haga clic en **Examinar...** para seleccionar un nombre de archivo y una ubicación.

Definición de valores de análisis de tamaño de efecto de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados binarios** > **Tamaño de efecto calculado previamente...**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis binario**, pulse **Análisis**.
3. Seleccione y defina los valores de análisis adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis binario: Inferencia

El diálogo **Inferencia** proporciona valores para especificar los métodos de estimación para el metaanálisis con resultados binarios cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Estimador
Proporciona los valores para especificar el estimador.

Notas:

- Cuando se especifican valores de Recorte y relleno, también controla el estimador que se utiliza mediante la agrupación en el análisis de recorte y relleno.
- Cuando se especifican valores de sesgo, también controla el estimador que utiliza la prueba basada en regresión.

Máxima verosimilitud restringida (REML)
El valor predeterminado aplica el método iterativo y calcula el estimador de máxima verosimilitud restringida.

Máxima verosimilitud (ML)
Aplica el método iterativo y calcula el estimador de máxima verosimilitud.

Bayes empírico
Aplica el método iterativo y calcula el estimador de Bayes empírico.

Hedges
Aplica el método no iterativo y calcula el estimador de Hedges.

Hunter-Schmidt
Aplica el método no iterativo y calcula el estimador de Hunter-Schmidt.

DerSimonian-Laird
Aplica el método no iterativo y calcula el estimador de DerSimonian-Laird.

Sidik-Jonkman
Aplica el método no iterativo y calcula el estimador de Sidik-Jonkman.

Ajuste del error estándar
Proporciona los valores que controlan si se aplica el ajuste de error estándar de Knapp-Hartung.

Notas:

- Cuando se especifican valores de Recorte y relleno, también controla el ajuste del error estándar que se utiliza mediante la agrupación en el análisis de recorte y relleno.
- Cuando se especifican valores de sesgo, también controla el ajuste del error estándar que utiliza la prueba basada en regresión.

Sin ajuste
El valor predeterminado no aplica el ajuste.

Aplicar el ajuste de Knapp-Hartung
Aplica el método de ajuste de Knapp-Hartung.

Aplicar el ajuste de Knapp-Hartung truncado
Aplica el método de ajuste de Knapp-Hartung y trunca el valor si es menor que 1 al estimar la matriz de varianza-covarianza.

Definición de valores de inferencia de tamaño de efecto de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados binarios** > **Tamaño de efecto calculado previamente...**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis binario**, pulse **Inferencia**.
3. Seleccione y defina los valores de inferencia adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis binario: Contraste

El diálogo **Contraste** proporciona valores para controlar la prueba de contraste para el metaanálisis con resultados binarios cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Variables
La lista muestra todas las variables de conjunto de datos disponibles. Seleccione variables de la lista y muévalas a la lista **Contraste(s)**.

Contraste(s)
La lista identifica los coeficientes que se almacenan como variables en el conjunto de datos activo. Se permiten varias variables. Las variables de cadena no están soportadas.

Valores de coeficiente de entrada de usuario
Proporciona valores para especificar los coeficientes de un contraste suministrado por el usuario. Solo se permiten valores numéricos. Para formular una prueba de contraste válida, el número de valores especificados debe coincidir con el de los estudios válidos.

Visualizar estadísticas exponenciadas
Controla la inclusión de los estadísticos en formato exponencial. Cuando se selecciona el valor, los resultados incluyen los estadísticos en formato exponencial, incluido el tamaño de efecto exponencial y los límites de intervalo de confianza. El valor está disponible cuando el **Tamaño de efecto** se especifica como **Razón de probabilidad de registro**, **Razón de probabilidad de registro Peto** o **Razón de riesgo de registro**.

Definición de valores de contraste de tamaño de efecto de meta-análisis binario

1. En los menús seleccione:

Analizar > **Meta-análisis** > **Resultados binarios** > **Tamaño de efecto calculado previamente...**

2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis binario**, pulse **Contraste**.
3. Seleccione y defina los valores de contraste adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis binario: Sesgo

El diálogo **Sesgo** proporciona valores para habilitar el sesgo de publicación llevando a cabo pruebas basadas en regresión para el metaanálisis con resultados binarios cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Prueba basada en regresión Egger
La selección de este valor habilita el sesgo de publicación realizando la prueba basada en regresión Eggers.

Variables
La lista proporciona todas las variables de conjunto de datos disponibles.

Covariables
Las variables seleccionadas de la lista **Variables** se tratan como covariables. Se permiten varias covariables.

Factores
Las variables seleccionadas de la lista **Variables** se tratan como factores. Se permiten varios factores.

Incluir la intersección en la regresión
Controla el término de intersección en la prueba basada en regresión.

Incluir parámetro de dispersión en el modelo de efectos fijos
Controla el ajuste de modelo multiplicativo e introduce el parámetro de dispersión multiplicativo en el análisis. El valor solo está disponible cuando se selecciona un modelo de efectos fijos.

Estimar estadísticas en función de la distribución t
Controla la distribución utilizada en las pruebas basadas en regresión. El valor está habilitado de forma predeterminada, lo que estima las estadísticas basadas en la distribución *t*. Cuando el valor no está seleccionado, las estadísticas se estiman en función de la distribución normal.

Definición de sesgo de tamaño de efecto de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados binarios** > **Tamaño de efecto calculado previamente...**

2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis binario**, pulse **Sesgo**.
3. Seleccione y defina los valores de sesgo adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis binario: Recortar y rellenar

El diálogo **Recorte y relleno** proporciona valores para implementar el análisis de recorte y relleno del sesgo de publicación para el metaanálisis con resultados binarios cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Número estimado de estudios perdidos
Controla el análisis de recorte y relleno de sesgo de publicación. La selección de este valor habilita los demás valores del diálogo.

Lado para imputar estudios
Proporciona opciones para especificar el lado del gráfico de embudo en el que se imputan los estudios perdidos.

Determinado por las inclinaciones de prueba Egger
El valor predeterminado determina el lado en función de la inclinación estimada de la prueba de Egger.

Izquierda
Imputa el lado izquierdo del gráfico de embudo.

Derecha
Imputa el lado derecho del gráfico de embudo.

Método
Especifica el método para estimar el número de estudios perdidos.

Lineal
El valor predeterminado calcula el estimador lineal.

Ejecute
Calcula el estimador de ejecución.

Cuadrático
Calcula el estimador cuadrático.

Proceso de iteración
Proporciona valores para especificar el estimador de iteración y el ajuste del error estándar.

Modelo de efectos fijos
Cuando se selecciona, se utiliza un modelo de efectos fijos y las opciones de estimación de iteración y ajuste de error estándar no están disponibles.

Modelo de efectos aleatorios
Cuando se selecciona, se utiliza un modelo de efectos aleatorios y están disponibles los valores siguientes.

Estimador
Proporciona valores para especificar el estimador de iteración

Máxima verosimilitud restringida (REML)
El valor predeterminado aplica el método iterativo y calcula el estimador de máxima verosimilitud restringida.

Máxima verosimilitud (ML)
Aplica el método iterativo y calcula el estimador de máxima verosimilitud.

Bayes empírico
Aplica el método iterativo y calcula el estimador de Bayes empírico.

Hedges
Aplica el método no iterativo y calcula el estimador de Hedges.

Hunter-Schmidt
Aplica el método no iterativo y calcula el estimador de Hunter-Schmidt.

DerSimonian-Laird
Aplica el método no iterativo y calcula el estimador de DerSimonian-Laird.

Sidik-Jonkman
Aplica el método no iterativo y calcula el estimador de Sidik-Jonkman.

Ajuste del error estándar
Proporciona valores para controlar si se aplica o no el ajuste de error estándar de Knapp-Hartung a las iteraciones del algoritmo de recorte y relleno.

Sin ajuste
El valor predeterminado no aplica el ajuste.

Aplicar el ajuste de Knapp-Hartung
Aplica el método de ajuste de Knapp-Hartung.

Aplicar el ajuste de Knapp-Hartung truncado
Aplica el método de ajuste de Knapp-Hartung y trunca el valor si es menor que 1 al estimar la matriz de varianza-covarianza.

Definición de recorte y relleno de tamaño de efecto de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados binarios** > **Tamaño de efecto calculado previamente...**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis binario**, pulse **Recortar y rellenar**.
3. Seleccione y defina los valores adecuados de recorte y relleno.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis binario: Imprimir

El diálogo **Imprimir** proporciona valores para controlar los resultados de tabla para el metaanálisis con resultados binarios cuando se proporcionan los datos de tamaño de efecto precalculados en el conjunto de datos activo.

Homogeneidad/Heterogeneidad
Proporciona valores para controlar la homogeneidad y la heterogeneidad.

Prueba de homogeneidad
Cuando se selecciona se proporciona la prueba de homogeneidad correspondiente en los resultados.

Medidas de heterogeneidad
Cuando se selecciona, se proporcionan las medidas de heterogeneidad en los resultados.

Tamaños de efecto
Proporciona los valores de tamaño de efecto siguientes.

Estudios individuales
Controla la visualización de los estudios individuales. Cuando se selecciona, se proporcionan los resultados correspondientes.

Tamaños de efecto acumulados
Controla la visualización del análisis acumulado. Cuando se selecciona, se proporcionan los resultados correspondientes. El valor solo está disponible cuando se selecciona una variable de **Análisis acumulado** en el diálogo **Análisis**.

Intervalo de predicción en el modelo de efectos aleatorios
Controla la visualización del intervalo de predicción. Cuando se selecciona, se proporcionan los resultados correspondientes. El valor solo está disponible cuando se especifica un modelo de efectos aleatorios.

Visualizar estadísticas exponenciadas
Controla la inclusión de los estadísticos en formato exponencial. Cuando se selecciona el valor, se generan los estadísticos en formato exponencial, incluido el tamaño de efecto exponencial y los límites de intervalo de confianza. El proceso se aplica a las tablas de predicción y estimación de tamaño de efecto opcionales y necesarias. El valor está disponible cuando el **Tamaño de efecto** se especifica como **Razón de probabilidad de registro**, **Razón de probabilidad de registro Peto** o **Razón de riesgo de registro**.

Definición de valores de impresión de tamaño de efecto de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados binarios** > **Tamaño de efecto calculado previamente...**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis binario**, pulse **Imprimir**.
3. Seleccione y defina los valores de impresión adecuados.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis binario: Guardar

El cuadro de diálogo **Guardar** proporciona valores para guardar las estadísticas estimadas en el conjunto de datos activo para el meta-análisis con resultados binarios cuando los datos de tamaño de efecto calculados previamente se proporcionan en el conjunto de datos activo.

Estudios individuales
Proporciona opciones para guardar el tamaño de efecto estimado.

Tamaño de efecto individual (formato exponenciado)
Guarda el tamaño de efecto estimado en formato exponencial.

Error estándar
Guarda el error estándar estimado del tamaño de efecto.

Límite inferior de intervalo de confianza
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto.

Límite superior de intervalo de confianza
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto.

Límite inferior de intervalo de confianza (formato exponenciado)
Guarda el límite inferior de intervalo de confianza estimado del tamaño de efecto en formato exponencial.

Límite superior de intervalo de confianza (formato exponenciado)
Guarda el límite superior de intervalo de confianza estimado del tamaño de efecto en formato exponencial.

Valor P
Guarda el valor *p* estimado del tamaño de efecto.

Ponderación de estudio
Guarda la ponderación de estudio estimada.

Porcentaje de ponderación de estudios
Guarda la ponderación de estudio normalizada como un porcentaje.

Definición de valores de guardado de tamaño de efecto de meta-análisis binario

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Resultados binarios** > **Tamaño de efecto calculado previamente...**
2. En el cuadro de diálogo **Tamaño de efecto de meta-análisis binario**, pulse **Guardar**.
3. Seleccione y defina los valores de guardado adecuados de estadísticas estimadas.
4. Pulse **Continuar**.

Tamaño de efecto de meta-análisis binario: Gráfico

El diálogo **Gráfico** proporciona valores para los siguientes tipos de gráfico:

- “Gráfico de bosque” en la página 78
- “Gráfico de bosque acumulado” en la página 79
- “Gráfico de burbujas” en la página 80
- “Gráfico de embudo” en la página 81
- “Gráfico de Galbraith” en la página 82
- “Gráfico de L'Abb ´e” en la página 83

Metaanálisis Regresión

El procedimiento Metaanálisis Regresión realiza un análisis de metarregresión.

Ejemplo

En el pasado, se llevaron a cabo varios estudios de investigación para investigar un medicamento novedoso pero controvertido para tratar la diabetes de tipo II. Se afirmaba que el medicamento oral reducía el nivel de glucosa en sangre después de las comidas. Se recopilaron datos de diferentes fuentes de investigación de 1979 a 1986.

Una de las directoras de la investigación pretendía extraer la inferencia estadística sobre el efecto del medicamento oral. Debido al hecho de que se generaron datos a partir de diferentes estudios, propuso la idea de sintetizar los resultados de todos los estudios para conocer el efecto global e identificar las fuentes subyacentes de variación en los resultados.

Estadísticos

Nivel de confianza, método iterativo, subdivisión de pasos, tolerancia de convergencia, medias de muestra, varianza de muestra, desviación estándar, tamaño de efecto estimado, método de estimación, prueba basada en regresión, modelo de efectos aleatorios, modelo de efectos fijos, parámetro de dispersión, estimador de máxima verosimilitud restringida, estimador de Bayes empírico, estimador de Hedges, estimador de Hunter-Schmidt, estimador de DerSimonian-Laird, estimador de Sidik-Jonkman, ajuste de error estándar de Knapp-Hartung, ajuste de error estándar de Knapp-Hartung truncado, coeficientes, prueba de coeficientes de modelo, estadísticos exponenciados.

Obtención de un análisis de Metaanálisis Regresión

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Meta-regresión**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione una variable **Tamaño de efecto** dependiente individual que denote el tamaño de efecto. La variable seleccionada debe ser numérica (no se admiten variables de cadena).

3. Seleccione uno de los valores siguientes y, a continuación, seleccione una variable numérica individual correspondiente:

   Error estándar
   Seleccione una variable que especifique el error estándar que se convierte en la ponderación. Ésta es el ajuste predeterminado.

   Varianza
   Seleccione una variable que especifique la varianza que se convierte en la ponderación.

   Ponderación
   Seleccione una variable que especifique la ponderación.

4. Si lo desea, añada variables de factor a la lista **Factores**. Para cada variable de factor seleccionada, se puede especificar una variable opcional para designar **Últimas categorías** personalizadas. Puede pulsar **Restablecer** para restaurar los valores de **Últimas categorías** a los valores predeterminados.

   Nota: Cuando no existan casos que coincidan con los valores de **Últimas categorías** especificados, los últimos valores que se hayan producido se tratarán como las últimas categorías.

5. Si lo desea, seleccione variables covariables numéricas.

6. Si lo desea, seleccione un valor de **Modelo**.

   Efectos aleatorios
   El valor predeterminado crea el modelo de efectos aleatorios.

   Efectos fijos
   Crea el modelo de efectos fijos. De forma opcional, puede seleccionar el valor **Incluir parámetro de dispersión**.

7. Si lo desea, puede:

   - Haz clic en **Criterios...** para especificar los criterios generales.
   - Pulse **Inferencia** para especificar los métodos de estimación.

- Pulse **Imprimir** para controlar los resultados de la tabla.
- Pulse **Guardar** para predecir y guardar los estadísticos estimados en el conjunto de datos activo.
- Pulse **Gráfico** para especificar qué gráficos incluir en la salida.

8. Pulse **Aceptar**.

Metaanálisis Regresión: Criterios

El diálogo **Criterios** proporciona valores para especificar los criterios para el metaanálisis regresión.

Intervalo de confianza
El valor opcional especifica el nivel de confianza. El valor debe estar comprendido por un valor numérico entre 0 y 100. El valor predeterminado es 95.

Valores perdidos de usuario
Los valores opcionales controlan cómo se tratan los valores perdidos de usuario.

Excluir
El valor predeterminado trata los valores perdidos de usuario como válidos.

Incluir
No tiene en cuenta las designaciones de valor perdido y trata los valores perdidos de usuario como válidos.

Iteraciones

Número máximo de iteraciones
El valor opcional especifica el número máximo de iteraciones en los métodos iterativos. El valor está disponible cuando se utilizan métodos iterativos. El valor debe ser un entero positivo único. El valor predeterminado es 100. El valor 0 indica que no se realizan iteraciones.

Máximo de subdivisión por pasos
El valor opcional especifica la máxima subdivisión por pasos en los métodos iterativos. El valor está disponible cuando se utilizan métodos iterativos. El valor debe ser un entero positivo único. El valor predeterminado es 5. Un valor de 0 significa que no se aplica la subdivisión por pasos.

Convergencia
El valor opcional especifica la tolerancia de convergencia. El valor debe ser un valor positivo único. El valor está disponible cuando se utilizan métodos iterativos. El valor predeterminado es 1E-6.

Definición de criterios del metaanálisis Regresión

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Meta-regresión**

2. En el diálogo **Metaanálisis Regresión**, pulse **Criterios**.
3. Seleccione y defina los valores de criterios adecuados.
4. Pulse **Continuar**.

Metaanálisis Regresión: Inferencia

El diálogo **Inferencia** proporciona valores para especificar los métodos de estimación para el metaanálisis regresión.

Incluir la intersección en la regresión
Este valor está habilitado de forma predeterminada.

Estimar estadísticas en función de la distribución t
Controla la distribución utilizada en las pruebas basadas en regresión. El valor está habilitado de forma predeterminada, lo que estima las estadísticas basadas en la distribución *t*. Cuando el valor no está seleccionado, las estadísticas se estiman en función de la distribución normal.

Estimador
Proporciona los valores para especificar el estimador.

Máxima verosimilitud restringida (REML)
El valor predeterminado aplica el método iterativo y calcula el estimador de máxima verosimilitud restringida.

Máxima verosimilitud (ML)
Aplica el método iterativo y calcula el estimador de máxima verosimilitud.

Bayes empírico
Aplica el método iterativo y calcula el estimador de Bayes empírico.

Hedges
Aplica el método no iterativo y calcula el estimador de Hedges.

Hunter-Schmidt
Aplica el método no iterativo y calcula el estimador de Hunter-Schmidt.

DerSimonian-Laird
Aplica el método no iterativo y calcula el estimador de DerSimonian-Laird.

Sidik-Jonkman
Aplica el método no iterativo y calcula el estimador de Sidik-Jonkman.

Ajuste del error estándar
Proporciona los valores que controlan si se aplica el ajuste de error estándar de Knapp-Hartung.

Sin ajuste
El valor predeterminado no aplica el ajuste.

Aplicar el ajuste de Knapp-Hartung
Aplica el método de ajuste de Knapp-Hartung.

Aplicar el ajuste de Knapp-Hartung truncado
Aplica el método de ajuste de Knapp-Hartung y trunca el valor si es menor que 1 al estimar la matriz de varianza-covarianza.

Definición de valores de inferencia del metaanálisis Regresión

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Meta-regresión**
2. En el diálogo **Metaanálisis Regresión**, pulse **Inferencia**.
3. Seleccione y defina los valores de inferencia adecuados.
4. Pulse **Continuar**.

Metaanálisis Regresión: Imprimir

El diálogo **Imprimir** proporciona valores para controlar los resultados de tabla para el metaanálisis regresión.

Prueba de coeficiente de modelo
Controla la prueba de coeficiente de modelo. El valor no está especificado de forma predeterminada, lo cual suprime la prueba. Cuando se especifica el valor, se proporciona la prueba en los resultados.

Visualizar estadísticas exponenciadas
Controla las estimaciones de parámetros. El valor no está especificado de forma predeterminada, lo cual suprime las estimaciones de parámetros. Cuando se especifica el valor, se proporcionan las estimaciones de parámetros en los resultados.

Definición de valores de impresión del metaanálisis Regresión

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Meta-regresión**
2. En el diálogo **Metaanálisis Regresión**, pulse **Imprimir**.
3. Seleccione y defina los valores de impresión adecuados.

4. Pulse **Continuar**.

Metaanálisis Regresión: Guardar

El diálogo **Guardar** proporciona valores para guardar los estadísticos estimados en el conjunto de datos activo para el metaanálisis regresión.

Estudios individuales
Proporciona opciones para guardar el tamaño de efecto estimado.

Valores pronosticados
Predice y guarda los valores ajustados.

Error estándar de valores pronosticados
Predice y guarda el error estándar de los valores ajustados.

Límite inferior de intervalo de confianza
Guarda el límite inferior del intervalo de confianza estimado de los valores ajustados.

Límite superior de intervalo de confianza
Guarda el límite superior del intervalo de confianza estimado de los valores ajustados.

Residuos
Predice y guarda los residuos.

Error estándar de residuos
Predice y guarda el error estándar de los residuos.

Influencias
Predice y guarda las influencias.

Estadísticos acumulados
Proporciona opciones para guardar prediciones lineales.

Nota: Los valores de estadísticos acumulados no se tienen en cuenta cuando se selecciona **Efectos fijos** como valor del **Modelo**.

Prediciones lineales fijas
Predice y guarda la predicción lineal.

Error estándar de prediciones lineales fijas
Predice y guarda el error estándar de la predicción lineal.

Mejores predicciones insesgadas lineales
Predice y guarda la mejor predicción insesgada lineal de los efectos aleatorios.

Error estándar de BLUP
Predice y guarda el error estándar de la mejor predicción insesgada lineal de los efectos aleatorios.

Definición de valores de guardado del metaanálisis Regresión

1. En los menús seleccione:

   Analizar > **Meta-análisis** > **Meta-regresión**
2. En el diálogo **Metaanálisis Regresión**, pulse **Guardar**.
3. Seleccione y defina los valores de guardado adecuados.
4. Pulse **Continuar**.

Metaanálisis Regresión: Gráfico

El diálogo **Gráfico** proporciona valores para definir gráficos de burbujas. Para obtener más información, consulte “Gráfico de burbujas” en la página 80.

Opciones de trazado de meta-análisis

Las siguientes opciones de gráficos están disponibles para los procedimientos de Meta-análisis y se puede acceder a ellas en los cuadros de diálogo de **Gráfico**.

Gráfico de bosque

La pestaña **Gráfico forestal** proporciona valores para controlar los gráficos forestales que se muestran en la salida para el meta-análisis con resultados continuos y binarios cuando los datos en bruto y los datos de tamaño de efecto calculados previamente se proporcionan en el conjunto de datos activo.

Gráfico de bosque
El valor opcional habilita e inhabilita los resultados del gráfico forestal. De forma predeterminada, el valor está inhabilitado.

Columnas de visualización
Los valores opcionales especifican las estadísticas estimadas que se tabulan.

Tamaño de efecto
Incluye el tamaño de efecto estimado.

Error estándar
Incluye el error estándar del tamaño de efecto estimado.

Límites de intervalo de confianza
Incluye los límites superiores e inferiores del intervalo de confianza del tamaño de efecto estimado.

Valor P
Incluye el valor *p* del tamaño de efecto estimado para cada estudio individual.

Ponderación
Incluye la ponderación de los estudios primarios.

Visualizar formato exponenciado

Nota: Este ajuste se aplica solo al meta-análisis con resultados binarios.

Controla la columna de gráfico y la visualización del tamaño de efecto y el intervalo de confianza. Cuando el valor está seleccionado, representa la razón de probabilidad, la razón de probabilidad de Peto o la proporción de riesgo, y tabula las estadísticas en formato exponencial. Cuando el valor no está seleccionado, representa la razón de probabilidad de registro, la razón de probabilidad de registro Peto o la razón de riesgo de registro, y tabula los estadísticos transformados logarítmicamente.

Variables
Lista las variables de conjunto de datos disponibles.

Columnas adicionales
Si lo desea, seleccione las variables que se mostrarán como columnas adicionales. El orden de las variables determina el orden de visualización de las columnas.

Ordenar
Opcionalmente, especifique una variable sobre la que se ordene el gráfico forestal. Cuando se especifica una variable, el gráfico forestal se clasifica en orden **Ascendente** (el valor predeterminado). Cuando se selecciona **Descendente**, el gráfico forestal se clasifica en orden descendente.

Posición de columna de gráfico
Controla la ubicación de la columna de gráfico.

Derecha
El valor predeterminado coloca la columna de gráfico a la derecha de las demás columnas de la tabla.

Izquierda
Coloca la columna de gráfico a la izquierda de las demás columnas de la tabla.

Líneas de referencia
Controla las líneas de referencia que se añaden al gráfico forestal. **Tamaño de efecto global** añade la línea que indica el tamaño de efecto global estimado. **Tamaño de efecto nulo** añade la línea que indica el tamaño de efecto nulo.

Anotaciones
Controla la visualización de las anotaciones.

Homogeneidad
Imprime los estadísticos de prueba de homogeneidad.

Heterogeneidad
Imprime los estadísticos de prueba de heterogeneidad.

Prueba
Imprime la prueba del tamaño de efecto global y la prueba de homogeneidad entre subgrupos en presencia del análisis de subgrupos.

Rango de recorte
Especifica el rango de recorte. Cuando se selecciona el ajuste, se deben especificar dos valores numéricos y satisfacer `Upper bound > Lower bound`.

Para meta-análisis con resultados binarios, cuando el ajuste de **Tamaño del Efecto** se establece en **Diferencia de Riesgo**, también debe satisfacer `Upper bound ≤ 1` y `Lower bound ≥ −1`. Cuando se selecciona el ajuste **Mostrar forma exponenciada**, también debe satisfacer `Lower bound ≥ 0`.

Definición de valores de gráfico de bosque

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados continuos o resultados binarios** > **Datos en bruto...** o **Tamaño del efecto precalculado...**
2. En el diálogo **Meta-análisis**, pulse **Gráfico** y, a continuación, pulse la pestaña **Gráfico forestal**.
3. Seleccione y defina los valores de gráficos forestales adecuados.
4. Pulse **Continuar**.

Gráfico de bosque acumulado

La pestaña **Gráfico de bosque acumulado** proporciona valores para controlar los gráficos de bosque acumulado que se muestran en los resultados del metaanálisis con resultados continuos y binarios cuando se proporcionan datos en bruto y datos con tamaño de efecto precalculado en el conjunto de datos activo.

Gráfico de bosque acumulado
El valor opcional habilita e inhabilita los resultados del gráfico forestal. De forma predeterminada, el valor está inhabilitado.

Columnas de visualización
Los valores opcionales especifican las estadísticas estimadas que se tabulan.

Tamaño de efecto acumulado
Incluye el tamaño de efecto global acumulado.

Error estándar
Incluye el error estándar del tamaño de efecto global acumulado.

Límites de intervalo de confianza
Incluye los límites inferior y superior del intervalo de confianza del tamaño de efecto global acumulado.

Valor P
Incluye el valor *p* del tamaño de efecto global acumulado para cada estudio individual.

Visualizar formato exponenciado

Nota: Este ajuste se aplica solo al meta-análisis con resultados binarios.

Controla la columna de gráfico y la visualización del tamaño de efecto y el intervalo de confianza. Cuando el valor está seleccionado, representa la razón de probabilidad, la razón de probabilidad de Peto o la proporción de riesgo, y tabula las estadísticas en formato exponencial. Cuando no se habilita el valor, representa la razón de probabilidad logarítmica, la razón de probabilidad logarítmica de Peto o la proporción de riesgo logarítmica, y tabula los estadísticos de transformación logarítmica.

Variables
Lista las variables de conjunto de datos disponibles.

Columnas adicionales
Si lo desea, seleccione las variables que se mostrarán como columnas adicionales. El orden de las variables determina el orden de visualización de las columnas.

Posición de columna de gráfico
Controla la ubicación de la columna de gráfico.

Derecha
El valor predeterminado coloca la columna de gráfico a la derecha de las demás columnas de la tabla.

Izquierda
Coloca la columna de gráfico a la izquierda de las demás columnas de la tabla.

Rango de recorte
Especifica el rango de recorte. Cuando se selecciona el ajuste, se deben especificar dos valores numéricos y satisfacer `Upper bound > Lower bound`.

Para meta-análisis con resultados binarios, cuando el ajuste de **Tamaño del Efecto** se establece en **Diferencia de Riesgo**, también debe satisfacer `Upper bound ≤ 1` y `Lower bound ≥ −1`. Cuando se selecciona el ajuste **Mostrar forma exponenciada**, también debe satisfacer `Lower bound ≥ 0`.

Definición de los valores de gráficos forestales acumulados

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados continuos o resultados binarios** > **Datos en bruto...** o **Tamaño del efecto precalculado...**
2. En el diálogo **Meta-análisis**, pulse **Gráfico** y, a continuación, pulse la pestaña **Gráfico forestal acumulado**.
3. Seleccione y defina los valores de gráfico de bosque acumulado adecuados.
4. Pulse **Continuar**.

Gráfico de burbujas

La pestaña **Gráfico de burbujas** proporciona valores para controlar los gráficos de burbujas que se muestran en la salida para el meta-análisis con resultados continuos y binarios cuando los datos en bruto y los datos de tamaño de efecto calculados previamente se proporcionan en el conjunto de datos activo.

Nota: El gráfico de burbujas es el único tipo de gráfico disponible para el análisis de regresión de meta-análisis.

Gráfico de burbujas
El valor opcional habilita e inhabilita los resultados del gráfico de burbujas. De forma predeterminada, el valor está inhabilitado.

Nota: Este valor no está disponible para el análisis de regresión de meta-análisis.

Variables
Lista las variables de conjunto de datos disponibles.

Predictor(es)
Especifique variables que se tratan como predictores continuos en el eje x. Se permiten varias variables y cada una de ellas crea un gráfico de burbujas independiente.

Nota: Para el análisis de regresión de meta-análisis, este valor está etiquetado como **Predictor(es) de gráfico de burbujas** y permite seleccionar covariables para las que crear gráficos de burbujas.

Predictor(es) centrado(s) en la media
El valor opcional controla el predictor continuo en el eje x. El valor está inhabilitado de forma predeterminada, lo que no centra el predictor. Cuando el valor está habilitado, el predictor está centrado en la media.

Etiqueta
Opcionalmente, especifique una variable sobre la que se etiquete el gráfico de burbujas. La lista **Colocación** proporciona opciones para colocar etiquetas automáticamente (**Automático**) o a la **Derecha**, a la **Izquierda**, en la parte **Superior** y en la parte **Inferior**.

Extraer burbujas proporcionales a las ponderaciones
Controla cómo se extraen las burbujas en relación con las ponderaciones. El valor está habilitado de forma predeterminada, lo que extrae las burbujas proporcionales a las ponderaciones. Cuando el valor está inhabilitado, todas las burbujas se extraen al mismo tamaño.

Mostrar línea ajustada
Controla la visualización de la línea de regresión ajustada. El valor está habilitado de forma predeterminada, lo que añade la línea de regresión ajustada al gráfico de burbujas. Cuando el valor está inhabilitado, la línea de regresión ajustada no se añade al gráfico de burbujas.

Mostrar límites de intervalo de confianza
Controla la visualización de los límites de intervalo de confianza. El valor está habilitado de forma predeterminada, lo que añade los límites de intervalo de confianza al gráfico de burbujas. Cuando el valor está inhabilitado, los límites de intervalo de confianza no se añaden al gráfico de burbujas.

Rango de eje x
Especifica el rango de gráfico de eje x. Cuando el ajuste está activado, se deben especificar dos valores numéricos y satisfacer `Upper bound > Lower bound`. Tenga en cuenta que los valores especificados se aplican a todos los gráficos de burbujas.

Rango de eje y
Especifica el rango de representación del eje Y. Cuando el ajuste está activado, se deben especificar dos valores numéricos y satisfacer `Upper bound > Lower bound`. Tenga en cuenta que los valores especificados se aplican a todos los gráficos de burbujas.

Para meta-análisis con resultados binarios, cuando el ajuste de **Tamaño del Efecto** se establece en **Diferencia de Riesgo**, también debe satisfacer `Upper bound ≤ 1` y `Lower bound ≥ −1`.

Definición de los valores de gráficos de burbujas

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados continuos o resultados binarios** > **Datos en bruto...** o **Tamaño del efecto precalculado...**

   o

   Analizar > **Meta-análisis** > **Meta-regresión**
2. En el diálogo **Meta-análisis**, pulse **Gráfico** y, a continuación, pulse la pestaña **Gráfico de burbujas**.
3. Seleccione y defina los valores de gráficos de burbujas adecuados.
4. Pulse **Continuar**.

Gráfico de embudo

La pestaña **Gráfico de embudo** proporciona valores para controlar los gráficos de embudo que se muestran en los resultados del metaanálisis con resultados continuos y binarios cuando se proporcionan datos en bruto y datos con tamaño de efecto precalculado en el conjunto de datos activo.

Gráfico de embudo
El valor opcional habilita e inhabilita la salida del gráfico de embudo. De forma predeterminada, el valor está inhabilitado.

Valores de eje Y
Controla los valores del eje Y.

Error estándar
Representa el error estándar.

Error estándar inverso
Representa el error estándar inverso.

Varianza
Representa la varianza.

Varianza inversa
Representa la varianza inversa.

Rango de eje x
Especifica el rango de gráfico de eje x. Cuando el ajuste está activado, se deben especificar dos valores numéricos y satisfacer `Upper bound > Lower bound`. Tenga en cuenta que los valores especificados se aplican a todos los gráficos de embudo.

Para meta-análisis con resultados binarios, cuando el ajuste **del Tamaño del Efecto** se establece en **Diferencia de Riesgo**, también debe satisfacer `Upper bound ≤ 1` y `Lower bound ≥ −1`.

Rango de eje y
Especifica el rango de representación del eje Y. Cuando el ajuste está activado, se deben especificar dos valores numéricos y satisfacer `Upper bound > Lower bound ≥ 0`. Tenga en cuenta que los valores especificados se aplican a todos los gráficos de embudo.

Incluir estudios imputados con Recortar y rellenar
Controla si se incluyen o no los estudios imputados en el gráfico de embudo cuando el valor **Número estimado de estudios perdidos** está habilitado en el diálogo **Recortar y rellenar**. El valor está inhabilitado de forma predeterminada, lo que excluye los estudios imputados. Cuando se selecciona **Mostrar tamaño de efecto global de los estudios observados**, se añade una línea de referencia vertical que denota el tamaño de efecto global estimado. El valor tiene en cuenta tanto los estudios observados como los estudios imputados.

Variables
Lista las variables de conjunto de datos disponibles.

Etiqueta
De forma opcional, especifique una variable en base a la cual se etiqueta el gráfico de embudo. La lista **Colocación** proporciona opciones para colocar etiquetas automáticamente (**Automático**) o a la **Derecha**, a la **Izquierda**, en la parte **Superior** y en la parte **Inferior**.

Definición de valores de gráfico de embudo

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados continuos o resultados binarios** > **Datos en bruto...** o **Tamaño del efecto precalculado...**
2. En el diálogo **Metaanálisis**, pulse **Gráfico** y, a continuación, pulse la pestaña **Gráfico de embudo**.
3. Seleccione y defina los valores de gráfico de embudo adecuados.
4. Pulse **Continuar**.

Gráfico de Galbraith

La pestaña **Gráfico de Galbraith** proporciona valores para controlar los gráficos de Galbraith que se muestran en los resultados del metaanálisis con resultados continuos y binarios cuando se proporcionan datos en bruto y datos con tamaño de efecto precalculado en el conjunto de datos activo.

Gráfico de Galbraith
El valor opcional habilita e inhabilita los resultados del gráfico de burbujas. De forma predeterminada, el valor está inhabilitado.

Mostrar límites de intervalo de confianza
Controla la visualización de los límites de intervalo de confianza. El valor está habilitado de forma predeterminada, lo que añade los límites de intervalo de confianza al gráfico de burbujas. Cuando el valor está inhabilitado, los límites de intervalo de confianza no se añaden al gráfico de burbujas.

Rango de eje x
Especifica el rango de gráfico de eje x. Cuando el ajuste está activado, se deben especificar dos valores numéricos y satisfacer `Upper bound > Lower bound ≥ 0`. Tenga en cuenta que los valores especificados se aplican a todos los gráficos de Galbraith.

Rango de eje y
Especifica el rango de representación del eje Y. Cuando el ajuste está activado, se deben especificar dos valores numéricos y satisfacer `Upper bound > Lower bound`. Tenga en cuenta que los valores especificados se aplican a todos los gráficos de Galbraith.

Variables
Lista las variables de conjunto de datos disponibles.

Etiqueta
Opcionalmente, especifique una variable sobre la que se etiquete el gráfico de burbujas. La lista **Colocación** proporciona opciones para colocar etiquetas automáticamente (**Automático**) o a la **Derecha**, a la **Izquierda**, en la parte **Superior** y en la parte **Inferior**.

Definición de los valores de gráficos de Galbraith

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados continuos o resultados binarios** > **Datos en bruto...** o **Tamaño del efecto precalculado...**
2. En el diálogo **Meta-análisis**, pulse **Gráfico** y, a continuación, pulse la pestaña **Gráfico de Galbraith**.
3. Seleccione y defina los valores de gráfico de Galbraith adecuados.
4. Pulse **Continuar**.

Gráfico de L'Abb´e

La pestaña **Gráfico de L'Abb´e** proporciona valores para controlar los gráficos de L'Abb´e que se muestran en la salida para el meta-análisis con resultados continuos y binarios cuando los datos en bruto y los datos de tamaño de efecto calculados previamente se proporcionan en el conjunto de datos activo.

Gráfico de L'Abb´e
El valor opcional habilita e inhabilita los resultados del gráfico de L'Abb´e. De forma predeterminada, el valor está inhabilitado.

Extraer puntos proporcionales a las ponderaciones
Controla cómo se extraen los puntos en relación con las ponderaciones. El valor está habilitado de forma predeterminada, lo que extrae los puntos proporcionales a las ponderaciones. Cuando el valor está inhabilitado, todos los puntos se extraen al mismo tamaño.

Líneas de referencia

Tamaño de efecto global
Controla la visualización de la línea de referencia que representa el tamaño de efecto global estimado. El valor está inhabilitado de forma predeterminada, lo que suprime la línea de referencia. La línea de referencia se visualiza cuando al valor está habilitado.

Tamaño de efecto nulo
Controla la visualización de la línea de referencia que no representa ningún efecto. El valor está habilitado de forma predeterminada, lo que muestra la línea de referencia. La línea de referencia se suprime cuando el valor no está habilitado.

Rango de eje x
Especifica el rango de gráfico de eje x. Cuando el ajuste está activado, se deben especificar dos valores numéricos y satisfacer `Upper bound > Lower bound`. Tenga en cuenta que los valores especificados se aplican a todos los gráficos de Galbraith.

Para meta-análisis con resultados binarios, cuando el ajuste de **Tamaño del Efecto** se establece en **Diferencia de Riesgo**, también debe satisfacer `Lower bound ≥ 0` y `Upper bound ≤ 1`. Cuando el ajuste **del tamaño del efecto** se establece en **Log Risk Ratio**, debe satisfacer `Upper bound ≤ 0`.

Rango de eje y
Especifica el rango de representación del eje Y. Cuando el ajuste está activado, se deben especificar dos valores numéricos y satisfacer `Upper bound > Lower bound`. Tenga en cuenta que los valores especificados se aplican a todos los gráficos de Galbraith.

Para meta-análisis con resultados binarios, cuando el ajuste de **Tamaño del Efecto** se establece en **Diferencia de Riesgo**, también debe satisfacer `Lower bound ≥ 0` y `Upper bound ≤ 1`. Cuando el ajuste **del tamaño del efecto** se establece en **Log Risk Ratio**, debe satisfacer `Upper bound ≤ 0`.

Variables
Lista las variables de conjunto de datos disponibles.

Etiqueta
Opcionalmente, especifique una variable sobre la que se etiquete el gráfico de burbujas. La lista **Colocación** proporciona opciones para colocar etiquetas automáticamente (**Automático**) o a la **Derecha**, a la **Izquierda**, en la parte **Superior** y en la parte **Inferior**.

Definición de los valores de gráficos de L'Abb´e

1. En los menús seleccione:

   Analizar > **Metaanálisis** > **Resultados continuos o resultados binarios** > **Datos en bruto...**
2. En el diálogo **Meta-análisis**, pulse **Gráfico** y, a continuación, pulse la pestaña **Gráfico de L'Abb´e**.
3. Seleccione y defina los valores de gráficos de L'Abb´e adecuados.
4. Pulse **Continuar**.

Libro de códigos

El libro de códigos hace referencia a la información del diccionario, como nombres de variable, etiquetas de variables, etiquetas de valores o valores perdidos, y los estadísticos de resumen de todas o las variables especificadas y conjuntos de respuestas múltiples del conjunto de datos activo. Para variables nominales y ordinales y conjuntos de respuestas múltiples, los estadísticos de resumen incluyen recuentos y porcentajes. Para variables de escala, los estadísticos de resumen incluyen la media, desviación estándar y cuartiles.

Nota: el libro de códigos ignora el estado del archivo segmentado. Esto incluye los grupos de archivos segmentados para imputaciones múltiples de valores perdidos (disponible en la opción adicional Valores perdidos).

Para obtener un libro de códigos

1. En los menús seleccione:

   Analizar > **Informes** > **Libro de códigos**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Pulse en la pestaña Variables.
3. Seleccione una o más variables y/o conjuntos de respuestas múltiples.

Si lo desea, puede:

- Controlar la información de variable que aparece.
- Controlar los estadísticos que aparecen (o excluir todos los estadísticos de resumen).

- Controlar el orden en que aparecen las variables y los conjuntos de respuestas múltiples.
- Cambiar el nivel de medición de cualquier variable en la lista de origen para modificar los estadísticos de resumen que aparecen. Consulte el tema “Pestaña Estadísticos del libro de códigos” en la página 87 para obtener más información.

Cambio del nivel de medición

Puede cambiar temporalmente el nivel de medición de variables. (No puede modificar el nivel de medición de conjuntos de respuestas múltiples. Se tratarán siempre como nominales.)

1. En la lista de origen, pulse con el botón derecho del ratón en una variable.
2. Seleccione un nivel de medición del menú emergente.

Se modificará el nivel de medición temporalmente. En términos prácticos, esto sólo es útil para variables numéricas. El nivel de medición de las variables de cadena está restringido a nominal u ordinal, los cuales reciben el mismo tratamiento por parte del procedimiento del libro de códigos.

Pestaña Resultados de libro de códigos

La pestaña Resultados controla la información de la variable incluida para cada variable y los conjuntos de respuestas múltiples, el orden en que aparecerán las variables y los conjuntos de respuestas múltiples y el contenido de la tabla de información del archivo opcional.

Información de variable

Controla la información del diccionario que se muestra para cada variable.

Posición. Un entero que representa la posición de la variable en el orden de archivo. No está disponible para conjuntos de respuestas múltiples.

Etiqueta. La etiqueta descriptiva asociada con la variable o el conjunto de respuesta múltiple.

Tipo. Tipos de datos fundamentales. Puede ser *Numérico*, *Cadena* o *Conjunto de respuesta múltiple*.

Formato. El formato de visualización de la variable, como *A4*, *F8.2* o *DATE11*. No está disponible para conjuntos de respuestas múltiples.

Nivel de medición. Los valores posibles son *Nominal*, *Ordinal*, *Escala* y *Desconocido*. El valor que aparece es el nivel de medición guardado en el diccionario y no se ve afectado por ninguna sustitución de medición temporal especificada al modificar el nivel de medición en la lista de variables de origen de la pestaña Variables. No está disponible para conjuntos de respuestas múltiples.

Nota: el nivel de medición de las variables numéricas puede ser "desconocido" antes de la primera lectura de datos si el nivel de medición no se ha definido de forma explícita, como lecturas de datos de un origen externo o nuevas variables creadas. Consulte el tema para obtener más información.

Papel. Algunos cuadros de diálogo permiten preseleccionar variables para su análisis en función de papeles definidos.

Etiquetas de valor. Etiquetas descriptivas asociadas con valores de datos específicos.

- Si selecciona Recuento o Porcentaje en la pestaña Estadísticos, las etiquetas de valor definidas se incluyen en la distribución de los resultados incluso si no selecciona las etiquetas de valor aquí.
- Para los conjuntos de dicotomías múltiples, las "etiquetas de valor" son etiquetas de variable de las variables elementales en el conjunto o las etiquetas de valores contados, dependiendo de cómo se define el conjunto. Consulte el tema para obtener más información.

Valores perdidos. Valores perdidos del usuario. Si selecciona Recuento o Porcentaje en la pestaña Estadísticos, las etiquetas de valor definidas se incluyen en la distribución de los resultados incluso si no selecciona los valores perdidos aquí. No está disponible para conjuntos de respuestas múltiples.

Atributos personalizados. Atributos de variable personalizados definidos por el usuario. Los resultados incluyen los nombres y valores de cualquier atributo de variable personalizado asociado con cada variable. Para obtener más información, consulte el tema . No está disponible para conjuntos de respuestas múltiples.

Atributos reservados. Atributos de variable de sistema reservados. Puede mostrar atributos del sistema, pero no debe modificarlas. Los nombres de atributos del sistema empiezan por un signo de dólar ($). Los atributos que no son de visualización, con nombres que empiezan por "@" o "$@", no se incluyen. Los resultados incluyen los nombres y valores de cualquier atributo de sistema asociado con cada variable. No está disponible para conjuntos de respuestas múltiples.

Información de archivo

La tabla de información del archivo opcional puede incluir cualquiera de los atributos de archivos siguientes:

Nombre de archivo. Nombre del archivo de datos de IBM SPSS Estadística . Si el conjunto de datos no se ha guardado nunca en formato de IBM SPSS Estadística , no existe un nombre de archivo de datos. (Si no aparece un nombre de archivo en la barra de título de la ventana del Editor de datos, el conjunto de datos activo no tiene un nombre de archivo.)

Posición. Ubicación del directorio (carpeta) del archivo de datos de IBM SPSS Estadística . Si el conjunto de datos no se ha guardado nunca en formato de IBM SPSS Estadística , no existe una ubicación.

Número de casos. Número de casos en el conjunto de datos activo. Es el número total de casos, incluyendo los casos que se pueden excluir de los estadísticos de resumen por condiciones de filtro.

Etiqueta. Es la etiqueta de archivo (si existe) definida por el comando `FILE LABEL` .

Documentos. Texto del documento del archivo de datos.

Estado de ponderación. Si se encuentra activada la ponderación, aparece el nombre de la variable de ponderación. Consulte el tema para obtener más información.

Atributos personalizados. Atributos de archivos de datos personalizados definidos por el usuario. Atributos de archivos de datos definidos con el comando `DATAFILE ATTRIBUTE`.

Atributos reservados. Atributos de archivo de datos de sistema reservados. Puede mostrar atributos del sistema, pero no debe modificarlas. Los nombres de atributos del sistema empiezan por un signo de dólar ($). Los atributos que no son de visualización, con nombres que empiezan por "@" o "$@", no se incluyen. Los resultados incluyen los nombres y valores de los atributos del archivo de datos del sistema.

Orden de visualización de variables

Las siguientes alternativas están disponibles para controlar el orden en que aparecen las variables y los conjuntos de respuestas múltiples.

Alfabético. Orden alfabético por nombre de variable.

Archivo. El orden en que aparecen las variables en el conjunto de datos (el orden en que aparecen en el editor de datos). En orden ascendente, los conjuntos de respuestas múltiples aparecen en último lugar, después de todas las variables seleccionadas.

Nivel de medición. Ordenar por nivel de medición. nominal, ordinal, escala y desconocido. Los conjuntos de respuestas múltiples se consideran nominales.

Nota: el nivel de medición de las variables numéricas puede ser "desconocido" antes de la primera lectura de datos si el nivel de medición no se ha definido de forma explícita, como lecturas de datos de un origen externo o nuevas variables creadas.

Lista de variables. El orden en que aparecen las variables y conjuntos de respuestas múltiples en la lista de variables seleccionadas en la pestaña Variables.

Nombre de atributo personalizado. La lista de opciones de orden de clasificación también incluye los nombres de cualquier atributo de variables personalizadas definidas por el usuario. En orden ascendente, las variables que no tienen la opción de clasificación de atributos al principio, seguidas de las variables que tienen el atributo pero no los valores definidos del atributo, seguidas de las variables con valores definidos para el atributo en orden alfabético de los valores.

Número máximo de categorías

Si el resultado incluye etiquetas de valor, los recuentos o porcentajes de cada valor exclusivo, puede eliminar esta información de la tabla si el número de los valores excede el valor especificado. De forma

predeterminada, esta información se elimina si el número de valores exclusivos de la variable es superior a 200.

Pestaña Estadísticos del libro de códigos

La pestaña Estadísticos permite controlar los estadísticos de resumen que se incluyen en los resultados o suprimir la visualización de los estadísticos de resumen completamente.

Recuentos y porcentajes

Para las variables nominales y ordinales, conjuntos de respuestas múltiples y valores de etiquetas de variables de escala, los estadísticos disponibles son:

Recuento. El recuento o número de casos que tienen cada valor (o el rango de valores) de una variable.

Porcentaje. Porcentaje de casos que presenta un valor determinado.

Tendencia y dispersión centrales

Para las variables de escala, los estadísticos disponibles son:

Media. Medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.

Desviación estándar. Una medida de dispersión sobre la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación estándar de la media y el 95% queda entre dos desviaciones estándar. Por ejemplo, si la edad media es de 45 años, con una desviación estándar de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.

Cuartiles. Muestra los valores correspondientes a los percentiles 25, 50 y 75.

Nota: puede modificar de forma temporal el nivel de medición asociado con una variable (y por lo tanto, modificar los estadísticos de resumen de la variable) en la lista de variables de origen de la pestaña Variables.

Frecuencias

El procedimiento Frecuencias proporciona estadísticos y representaciones gráficas que resultan útiles para describir muchos tipos de variables. El procedimiento Frecuencias es un comienzo para empezar a consultar los datos.

Para los informes de frecuencias y los gráficos de barras, puede organizar los valores distintos en orden ascendente o descendente u ordenar las categorías por sus frecuencias. Es posible suprimir el informe de frecuencias cuando una variable posee muchos valores distintos. Puede etiquetar los gráficos con las frecuencias (la opción predeterminada) o con los porcentajes.

Ejemplo

¿Cuál es la distribución de los clientes de una empresa por tipo de industria? En los resultados podría observar que el 37,5% de sus clientes pertenece a agencias gubernamentales, el 24,9% a corporaciones, el 28,1% a instituciones académicas, y el 9,4% a la industria sanitaria. Con respecto a los datos continuos, cuantitativos, como los ingresos por ventas, podría comprobar que el promedio de ventas de productos es de 3.576 dólares con una desviación estándar de 1.078 dólares.

Estadísticos y gráficos

Frecuencias, porcentajes, porcentajes acumulados, media, mediana, modo, suma, desviación estándar, varianza, amplitud, valores mínimo y máximo, error estándar de la media, asimetría y curtosis (ambos con sus errores estándar), cuartiles, percentiles especificados por el usuario, gráficos de barras, gráficos circulares e histogramas.

Consideraciones sobre los datos

Datos

Utilice códigos numéricos o cadenas para codificar las variables categóricas (mediciones de nivel nominal u ordinal).

Supuestos

Las tabulaciones y los porcentajes proporcionan una descripción útil para los datos de cualquier distribución, especialmente para las variables con categorías ordenadas o desordenadas. Muchos de los estadísticos de resumen optativos, tales como la media y la desviación estándar, se basan en la teoría normal y son apropiados para las variables cuantitativas con distribuciones simétricas. Los estadísticos robustos, tales como la mediana, los cuartiles y los percentiles son apropiados para las variables cuantitativas que pueden o no cumplir el supuesto de normalidad.

Obtención de tablas de frecuencias

1. En los menús seleccione:

   Analizar > **Estadísticas descriptivas** > **Descriptivos...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables categóricas o cuantitativas.
3. Si lo desea, seleccione **Crear tablas de estilos APA** para crear tablas de salida que se adhieran a las directrices de estilo APA.
4. Si lo desea, puede:
   - Pulse **Estadísticos** para obtener estadísticos descriptivos para las variables cuantitativas.
   - Pulsar en **Gráficos** para obtener gráficos de barras, gráficos circulares e histogramas.
   - Pulsar en **Formato** para determinar el orden en el que se muestran los resultados.
   - Pulsar **Estilo** para especificar las condiciones para cambiar automáticamente las propiedades de las tablas dinámicas en función de condiciones específicas.
   - Pulsar **Programa de arranque** para derivar estimaciones robustas de errores estándar e intervalos de confianza para estimaciones como, por ejemplo, la media, mediana, proporción, razón de las ventajas, coeficiente de correlación o coeficiente de regresión. También se puede utilizar para crear contrastes de hipótesis.

Frecuencias: Estadísticos

Valores percentiles. Los valores de una variable cuantitativa que dividen los datos ordenados en grupos, de forma que un porcentaje de los casos se encuentre por encima y otro porcentaje se encuentre por debajo. Los cuartiles (los percentiles 25, 50 y 75) dividen las observaciones en cuatro grupos de igual tamaño. Si desea un número igual de grupos que no sea cuatro, seleccione **Puntos de corte para n grup os iguales.** También puede especificar percentiles individuales (por ejemplo, el percentil 95, el valor por debajo del cual se encuentran el 95% de las observaciones).

Tendencia central. Los estadísticos que describen la localización de la distribución, incluyen: Media, Mediana, Moda y Suma de todos los valores.

- *Media*. Medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.
- *Mediana*. El valor por encima y por debajo del cual se sitúan la mitad de los casos: el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).
- *Modo*. El valor que se produce con más frecuencia. Si varios valores comparten la mayor frecuencia de aparición, cada uno de ellos es un modo. El procedimiento de frecuencias devuelve sólo el modo más pequeño de los modos múltiples.
- *Suma*. Suma o total de los valores, a lo largo de todos los casos con valores no perdidos.

Dispersión. Los estadísticos que miden la cantidad de variación o de dispersión en los datos, incluyen: Desviación estándar, Varianza, Rango, Mínimo, Máximo y Error estándar de la media.

- *Desviación desviación*. Una medida de dispersión sobre la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación estándar de la media y el 95% queda entre dos desviaciones estándar. Por ejemplo, si la edad media es de 45 años, con una desviación estándar de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.
- *Varianza*. Medida de dispersión sobre la media, igual a la suma de las desviaciones al cuadrado de la media divida por el número de casos menos uno. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.
- *Rango*. Diferencia entre los valores mayor y menor de una variable numérica; el máximo menos el mínimo.
- *Mínimo*. El valor menor de una variable numérica.
- *máximo*. El valor mayor de una variable numérica.
- *S. E. media*. Medida de cuánto puede variar el valor de la media de una muestra a otra tomada de la misma distribución. Puede utilizarse para comparar de forma aproximada la media observada respecto a un valor hipotetizado (es decir, se puede concluir que los dos valores son distintos si la diferencia entre ellos, dividida por el error estándar, es menor que -2 o mayor que +2).
- *Currículum vitae*. La razón de la desviación estándar a la media. Los valores se expresan en porcentajes.

Distribución. Asimetría y curtosis son estadísticos que describen la forma y la simetría de la distribución. Estos estadísticos se muestran con sus errores estándar.

- *Asimetría*. Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de sesgo de 0. Una distribución con un sesgo positivo significativo tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error estándar se asume que indica una desviación de la simetría.
- *Curtosis*. Medida del grado en que hay valores atípicos. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que los datos muestran más valores atípico extremos que una distribución normal. La curtosis negativa indica que los datos presentan valores atípicos menos extremados que una distribución normal. Algunas veces se hace referencia a la definición de curtosis que se utiliza, donde el valor es 0 para una distribución normal, como curtosis de exceso. Algunos programas pueden notificar curtosis de tal manera que el valor es 3 para una distribución normal.

Los valores son puntos medios de grupos. Si los valores de los datos son puntos medios de grupos (por ejemplo, si las edades de todas las personas entre treinta y cuarenta años se codifican como 35), seleccione esta opción para estimar la mediana y los percentiles para los datos originales no agrupados.

Frecuencias: Gráficos

Nota: Los gráficos no se generan en la salida cuando **Realizar simulación de muestreo** está habilitado en el diálogo Simulación de muestreo.

Tipo de gráfico

Los gráficos circulares muestran la contribución de las partes a un todo. Cada porción de un gráfico circular corresponde a un grupo, definido por una única variable de agrupación. Los gráficos de barras muestran el recuento de cada valor o categoría distinta como una barra diferente, permitiendo comparar las categorías de forma visual. Los histogramas también cuentan con barras, pero se representan a lo largo de una escala de intervalos iguales. La altura de cada barra es el recuento de los valores que están dentro del intervalo para una variable cuantitativa. Los histogramas muestran la forma, el centro y la dispersión de la distribución. Una curva normal superpuesta en un histograma ayuda a juzgar si los datos están normalmente distribuidos.

Valores del gráfico

Para los gráficos de barras, el eje de escala se puede etiquetar mediante porcentajes o recuentos de frecuencia.

Frecuencias: Formato

Ordenar por. La tabla de frecuencias se puede organizar respecto a los valores actuales de los datos o respecto al recuento (frecuencia de aparición) de esos valores y la tabla puede organizarse en orden ascendente o descendente. Sin embargo, si solicita un histograma o percentiles, Frecuencias asumirá que la variable es cuantitativa y mostrará sus valores en orden ascendente.

Múltiples variables. Si desea generar tablas de estadísticos para múltiples variables, podrá mostrar todas las variables en una sola tabla (**Comparar variables**), o bien mostrar una tabla de estadísticos independiente para cada variable (**Organizar resultados según variables)**.

Suprimir tablas con varias categorías. Esta opción impide que se muestren tablas que contengan más valores que el número especificado.

Descriptivos

El procedimiento Descriptivos muestra estadísticos de resumen univariados para varias variables en una única tabla y calcula valores tipificados (puntuaciones *z*). Las variables se pueden ordenar por el tamaño de sus medias (en orden ascendente o descendente), alfabéticamente o por el orden en el que se seleccionen las variables (el valor predeterminado).

Cuando se guardan las puntuaciones *z*, éstas se añaden a los datos del Editor de datos y quedan disponibles para los gráficos, el listado de los datos y los análisis. Cuando las variables se registran en unidades diferentes (por ejemplo, producto interior bruto per cápita y porcentaje de alfabetización), una transformación de puntuación *z* pondrá las variables en una escala común para poder compararlas visualmente con más facilidad.

Ejemplo. Si cada caso de los datos contiene los totales de ventas diarias de cada vendedor (por ejemplo, una entrada para Bob, una para Kim y una para Brian) recogidas cada día durante varios meses, el procedimiento Descriptivos puede calcular la media diaria de ventas para cada vendedor y ordenar los resultados del promedio de ventas de mayor a menor.

Estadísticas. Tamaño de la muestra, media, mínimo, máximo, desviación estándar, varianza, rango, suma, error estándar de la media, curtosis y asimetría con sus errores estándar.

Descriptivos: Consideraciones sobres los datos

Datos. Utilice variables numéricas después de haberlas inspeccionado gráficamente para registrar errores, valores atípicos y anomalías de distribución. El procedimiento Descriptivos es muy eficaz para archivos grandes (de miles de casos).

Supuestos. La mayoría de los estadísticos disponibles (incluyendo las puntuaciones *z*) se basan en la teoría normal y son adecuados para variables cuantitativas (mediciones a nivel de razón o de intervalo) con distribuciones simétricas. Se deben evitar las variables con categorías no ordenadas o distribuciones asimétricas. La distribución de puntuaciones *z* tiene la misma forma que la de los datos originales; por tanto, el cálculo de puntuaciones *z* no es una solución para los datos con problemas.

Para obtener estadísticos descriptivos

1. En los menús seleccione:

   Analizar > **Estadísticas descriptivas** > **Descriptivos...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables.

Si lo desea, puede:

- Seleccionar **Guardar valores tipificados como variables** para guardar las puntuaciones *z* como nuevas variables.
- Pulsar en **Opciones** para seleccionar estadísticos opcionales y el orden de presentación.

Descriptivos: Opciones

Media y suma. Se muestra de forma predeterminada la media o promedio aritmético.

Dispersión. Los estadísticos que miden la dispersión o variación en los datos incluyen la desviación estándar, la varianza, el rango, el mínimo, el máximo y el error estándar de la media.

- *Desviación desviación*. Una medida de dispersión sobre la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación estándar de la media y el 95% queda entre dos desviaciones estándar. Por ejemplo, si la edad media es de 45 años, con una desviación estándar de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.
- *Varianza*. Medida de dispersión sobre la media, igual a la suma de las desviaciones al cuadrado de la media divida por el número de casos menos uno. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.
- *Rango*. Diferencia entre los valores mayor y menor de una variable numérica; el máximo menos el mínimo.
- *Mínimo*. El valor menor de una variable numérica.
- *máximo*. El valor mayor de una variable numérica.
- *Error estándar de la media*. Medida de cuánto puede variar el valor de la media de una muestra a otra tomada de la misma distribución. Puede utilizarse para comparar de forma aproximada la media observada respecto a un valor hipotetizado (es decir, se puede concluir que los dos valores son distintos si la diferencia entre ellos, dividida por el error estándar, es menor que -2 o mayor que +2).
- *Currículum vitae*. La razón de la desviación estándar a la media. Los valores se expresan en porcentajes.

Distribución. La curtosis y la asimetría son los estadísticos que caracterizan la forma y simetría de la distribución. Estos estadísticos se muestran con sus errores estándar.

- *Curtosis*. Medida del grado en que hay valores atípicos. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que los datos muestran más valores atípico extremos que una distribución normal. La curtosis negativa indica que los datos presentan valores atípicos menos extremados que una distribución normal. Algunas veces se hace referencia a la definición de curtosis que se utiliza, donde el valor es 0 para una distribución normal, como curtosis de exceso. Algunos programas pueden notificar curtosis de tal manera que el valor es 3 para una distribución normal.
- *Asimetría*. Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de sesgo de 0. Una distribución con un sesgo positivo significativo tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error estándar se asume que indica una desviación de la simetría.

Orden de presentación. De forma predeterminada, las variables se muestran en el orden en que se hayan seleccionado. Si lo desea, se pueden mostrar las variables alfabéticamente, por medias ascendentes o por medias descendentes.

Características adicionales del comando DESCRIPTIVES

La sintaxis de comandos también le permite:

- Guardar puntuaciones estandarizadas (puntuaciones *z*) para algunas, pero no todas las variables (con el subcomando VARIABLES).
- Especificar nombres para las variables nuevas que contienen puntuaciones estandarizadas (con el subcomando VARIABLES).
- Excluir del análisis los casos con valores perdidos para cualquier variable (con el subcomando MISSING).
- Clasificar las variables de la visualización según el valor de cualquier estadística, no sólo la media (con el subcomando SORT).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Percentiles

El procedimiento Percentiles muestra tablas de percentiles.

Valores de percentiles
Cuando selecciona **Valores percentiles**, de forma predeterminada se muestran los valores correspondientes a los percentiles 5, 10, 25, 50, 75, 90 y 95. Los valores de una variable cuantitativa que dividen los datos ordenados en grupos, de forma que un porcentaje de los casos se encuentre por encima y otro porcentaje se encuentre por debajo.

Cuartiles
Los cuartiles (los percentiles 25, 50 y 75) dividen las observaciones en cuatro grupos de igual tamaño.

Personalizado
Cuando seleccione **Personalizado**, especifique al menos un valor para ejecutar el análisis. Los valores de entrada deben ser un número comprendido entre 0 y 100. Utilice los botones **Añadir**, **Cambiar**y **Eliminar** para trabajar con los valores de la lista de valores percentiles.

Método de percentil
De forma predeterminada, el método HAVERAGE está seleccionado para calcular percentiles.

Obtención de tablas de percentiles

1. En los menús seleccione:

   Analizar > **Estadísticos descriptivos** > **Percentiles ..**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables numéricas y desplácelas al campo Variables.
3. Seleccione **Valores de percentil** para especificar valores para calcular el percentil.
4. Seleccione **Método de percentil** para seleccionar el método de cálculo de percentil.
5. Si lo desea, puede:
   - Pulse **Falta** para controlar el tratamiento de los valores perdidos.
   - Pulse **Arranque** para derivar estimaciones robustas de errores estándar e intervalos de confianza para estimaciones de percentiles.

Percentiles de valores perdidos

Valores perdidos
Controla el tratamiento de los valores perdidos.

Excluir casos según lista

Los casos con valores perdidos para cualquier variable especificada se excluyen de todos los análisis. Este es el valor predeterminado.

Por variable
Excluir casos con valores perdidos en una base variable por variable.

Explorar

El procedimiento Explorar genera estadísticos de resumen y representaciones gráficas, bien para todos los casos o bien de forma separada para grupos de casos. Existen numerosas razones para utilizar este procedimiento: para inspeccionar los datos, identificar valores atípicos, obtener descripciones, comprobar supuestos y caracterizar diferencias entre subpoblaciones (grupos de casos). La inspección de los datos puede mostrar que existen valores inusuales, valores extremos, discontinuidades en los datos u otras peculiaridades. La exploración de los datos puede ayudar a determinar si son adecuadas las técnicas estadísticas que está teniendo en consideración para el análisis de los datos. La exploración

puede indicar que necesita transformar los datos si la técnica necesita una distribución normal. O bien, el usuario puede decidir que necesita utilizar pruebas no paramétricas.

Ejemplo. Observe la distribución de los tiempos de aprendizaje de laberintos de una serie de ratas sometidas a cuatro programas de refuerzo diferentes. Para cada uno de los cuatro grupos, se puede observar si la distribución de tiempos es aproximadamente normal y si las cuatro varianzas son iguales. También se pueden identificar los casos con los cinco valores de tiempo mayores y los cinco menores. Los diagramas de caja y los gráficos de tallo y hojas resumen gráficamente la distribución del tiempo de aprendizaje de cada uno de los grupos.

Estadísticos y gráficos. Media, mediana, media recortada al 5%, error estándar, varianza, desviación estándar, mínimo, máximo, rango, rango intercuartil, asimetría y curtosis y sus errores estándar, intervalo de confianza para la media (y el nivel de confianza especificado), percentiles, estimador-M de Huber, estimador en onda de Andrews, estimador-M redescendente de Hampel, estimador biponderado de Tukey, cinco valores mayores y cinco menores, estadístico de Kolmogorov-Smirnov con el nivel de significación de Lilliefors para contrastar la normalidad y estadístico de Shapiro-Wilk. Diagramas de caja, gráficos de tallo y hojas, histogramas, diagramas de normalidad y diagramas de dispersión por nivel con pruebas de Levene y transformaciones.

Explorar: Consideraciones sobre los datos

Datos. El procedimiento Explorar se puede utilizar para las variables cuantitativas (nivel de medición de razón o de intervalo). Una variable de factor (utilizada para dividir los datos en grupos de casos) debe tener un número razonable de valores distintivos (categorías). Estos valores pueden ser de cadena corta o numéricos. La variable de etiquetas de caso, utilizada para etiquetar valores atípicos en los diagramas de caja, puede ser de cadena corta, de cadena larga (los 15 primeros bytes) o numérica.

Supuestos. La distribución de los datos no tiene que ser simétrica ni normal.

Para explorar los datos

1. En los menús seleccione:

   Analizar > **Estadísticos descriptivos** > **Explorar...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables dependientes.

Si lo desea, puede:

- Seleccionar una o más variables de factor, cuyos valores definirán grupos de casos.
- Seleccionar una variable de identificación para etiquetar los casos.
- Pulse en **Estadísticos** para obtener estimadores robustos, valores atípicos, percentiles y tablas de frecuencias.
- Pulse en **Gráficos** para obtener histogramas, pruebas y gráficos de probabilidad normal y diagramas de dispersión por nivel con estadísticos de Levene.
- Pulse **Opciones** para el tratamiento de los valores perdidos.

Explorar: Estadísticos

Descriptivos. De forma predeterminada se muestran estas medidas de dispersión y de tendencia central. Éstas últimas indican la localización de la distribución, e incluyen la media, la mediana y la media recortada al 5%. Las medidas de dispersión muestran la disimilaridad de los valores, incluyen: los errores estándar, la varianza, la desviación estándar, el mínimo, el máximo, el rango y el rango intercuartil.
Los estadísticos descriptivos también incluyen medidas de la forma de la distribución: la asimetría y la curtosis se muestran con sus errores estándar. También se muestra el intervalo de confianza a un nivel del 95%; aunque se puede especificar otro nivel.

Estimadores robustos centrales. Alternativas robustas a la mediana y a la media muestral para estimar la localización. Los estimadores que se calculan difieren en las ponderaciones que aplican a los casos.

Se muestran los siguientes: el estimador-M de Huber, el estimador en onda de Andrew, el estimador-M redescendente de Hampel y el estimador biponderado de Tukey.

Valores atípicos. Muestra los cinco valores mayores y los cinco menores con las etiquetas de caso.

Percentiles

El procedimiento Percentiles muestra tablas de percentiles. Al seleccionar **«Percentil»**, se muestran de forma predeterminada los valores correspondientes a los percentiles 5, 10, 25, 50, 75, 90 y 95. Los valores de una variable cuantitativa que dividen los datos ordenados en grupos, de forma que un porcentaje de los casos se encuentre por encima y otro porcentaje se encuentre por debajo.

Cuartiles

Los cuartiles (los percentiles 25, 50 y 75) dividen las observaciones en cuatro grupos de igual tamaño.

Personalizado

Cuando seleccione **Personalizado**, especifique al menos un valor para ejecutar el análisis. Los valores de entrada deben ser un número comprendido entre 0 y 100. Utilice los botones **Añadir**, **Cambiar**y **Eliminar** para trabajar con los valores de la lista de valores percentiles.

Método

De forma predeterminada, el método HAVERAGE está seleccionado para calcular percentiles.

Explorar: Gráficos

Diagramas de caja. Estas alternativas controlan la presentación de los diagramas de caja cuando existe más de una variable dependiente. **Niveles de los factores juntos** genera una presentación para cada variable dependiente. En cada una se muestran diagramas de caja para cada uno de los grupos definidos por una variable de factor. **Dependientes juntas** genera una presentación para cada grupo definido por una variable de factor. En cada una se muestran juntos los diagramas de caja de cada variable dependiente. Esta disposición es de gran utilidad cuando las variables representan una misma característica medida en momentos distintos.

Descriptivos. La sección Descriptivos permite seleccionar gráficos de tallo y hojas e histogramas.

Gráficos con pruebas de normalidad. Muestra los diagramas de probabilidad normal y de probabilidad sin tendencia. Se muestra el estadístico de Kolmogorov-Smirnov con un nivel de significación de Lilliefors para contrastar la normalidad. Si se especifican ponderaciones no enteras, el estadístico de Shapiro-Wilk se calcula cuando el tamaño de la muestra ponderada se encuentra entre 3 y 50. En el caso de que no haya ponderaciones o de que las ponderaciones sean enteras, la estadística se calcula cuando la muestra está en el rango 3 - 5000.

Dispersión por nivel con prueba de Levene. Controla la transformación de los datos para los diagramas de dispersión por nivel. Para todos los diagramas de dispersión por nivel se muestra la inclinación de la línea de regresión y las pruebas robustas de Levene sobre la homogeneidad de varianza. Si selecciona una transformación, las pruebas de Levene se basarán en los datos transformados. Si no selecciona ninguna variable de factor, no se generará ningún diagrama de dispersión por nivel. **Estimación de potencia** produce un gráfico de los logaritmos naturales de los rangos intercuartiles respecto a los logaritmos naturales de las medianas de todas las casillas, así como una estimación de la transformación de potencia necesaria para conseguir varianzas iguales en las casillas. Un diagrama de dispersión por nivel ayuda a determinar la potencia que precisa una transformación para estabilizar (igualar) las varianzas de los grupos. **Transformados** permite seleccionar una de las alternativas de potencia, quizás siguiendo las recomendaciones de la estimación de potencia, y genera gráficos de los datos transformados. Se trazan el rango intercuartil y la mediana de los datos transformados. **No transformados** genera gráficos de los datos en bruto. Es equivalente a una transformación con una potencia de 1.

Explorar: Transformaciones de potencia

A continuación aparecen las transformaciones de potencia para los diagramas de dispersión por nivel. Para transformar los datos, deberá seleccionar una potencia para la transformación. Puede elegir una de las siguientes alternativas:

- **Log natural.** Transformación de logaritmo natural. Este es el valor predeterminado.
- **1/raíz cuadrada.** Para cada valor de los datos se calcula el inverso de la raíz cuadrada.
- **Recíproco.** Se calcula el inverso de cada valor de los datos.
- **Raíz cuadrada.** Se calcula la raíz cuadrada de cada valor de los datos.
- **Cuadrado.** Se calcula el cuadrado de cada valor de los datos.
- **Cubo.** Se calcula el cubo de cada valor de los datos.

Explorar: Opciones

Valores perdidos. Controla el tratamiento de los valores perdidos.

- **Excluir casos según lista.** Los casos con valores perdidos para cualquier variable de factor o variable dependiente se excluyen de todos los análisis. Este es el valor predeterminado.
- **Excluir casos según pareja.** Los casos con valores no perdidos para las variables de un grupo (casilla) se incluyen en el análisis de ese grupo. El caso puede tener valores perdidos para las variables utilizadas en otros grupos.
- **Mostrar los valores.** Los valores perdidos para las variables de factor se tratan como una categoría diferente. Todos los resultados se generan para esta categoría adicional. Las tablas de frecuencias incluyen categorías para los valores perdidos. Los valores perdidos para una variable de factor se incluyen pero se etiquetan como perdidos.

Características adicionales del comando EXAMINE

El procedimiento Explorar utiliza la sintaxis de comandos EXAMINE. La sintaxis de comandos también le permite:

- Solicitar los gráficos y resultados totales además de los gráficos y los resultados para los grupos definidos por las variables de factor (con el subcomando TOTAL).
- Especificar una escala común para un grupo de diagramas de caja (con el subcomando SCALE).
- Especificar interacciones de variables de factor (con el subcomando VARIABLES).
- Especificar percentiles distintos de los percentiles predeterminados (con el subcomando PERCENTILES).
- Calcular percentiles respecto a cualquiera de los cinco métodos (con el subcomando PERCENTILES).
- Especificar una transformación de potencia para gráficos de dispersión versus nivel (con el subcomando PLOT).
- Especificar el número de valores extremos que se van a mostrar (mediante el subcomando STATISTICS).
- Especificar parámetros para los estimadores M, los estimadores robustos de ubicación (mediante el subcomando MESTIMATORS).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Tablas cruzadas

El procedimiento Tablas cruzadas crea tablas bidimensionales y multidimensionales y, además, proporciona una serie de pruebas y medidas de asociación para las tablas bidimensionales. La estructura de la tabla y el hecho de que las categorías estén ordenadas o no determinan las pruebas o medidas que se utilizaban.

Con la excepción de los coeficientes de gamas parciales, los estadísticos de Tablas cruzadas y las medidas de asociación se calculan por separado para cada tabla bidireccional. Si especifica una fila, una columna y un factor de capa (variable de control), el procedimiento Tablas cruzadas crea un panel de medidas y estadísticos asociados para cada valor del factor de capa (o una combinación de valores para dos o más variables de control). Por ejemplo, si *sexo* es un factor de capa para una tabla de *casado* (sí, no) en función de *vida* (vida emocionante, rutinaria o aburrida), los resultados para una tabla bidimensional

para las mujeres se calculan de forma independiente de los resultados de los hombres y se imprimen en paneles uno detrás del otro.

Ejemplo. ¿Es más probable que los clientes de las empresas pequeñas sean más rentables en la venta de servicios (por ejemplo, formación y asesoramiento) que los clientes de las empresas grandes? A partir de una tabla cruzada podría deducir que la prestación de servicios a la mayoría de las empresas pequeñas (con menos de 500 empleados) produce considerables beneficios, mientras que con la mayoría de las empresas de gran tamaño (con más de 2.500 empleados), los beneficios obtenidos son mucho menores.

Estadísticos y medidas de asociación. Chi-cuadrado de Pearson, chi-cuadrado de la razón de verosimilitud, prueba de asociación lineal por lineal, prueba exacta de Fisher, chi-cuadrado corregido de Yates, *r* de Pearson, rho de Spearman, coeficiente de contingencia, phi, *V* de Cramér, lambdas simétricas y asimétricas, tau de Kruskal y Goodman, coeficiente de incertidumbre, gamma, *d* de Somers, tau-*b* de Kendall, tau-*c* de Kendall, coeficiente eta, kappa de Cohen, estimación de riesgo relativo, razón de las ventajas, prueba de McNemar y estadísticos de Cochran y Mantel-Haenszel.

Tablas cruzadas: Consideraciones sobre los datos

Datos. Para definir las categorías de cada variable de la tabla, utilice valores de una variable numérica o de cadena. Por ejemplo, para *sexo*, podría codificar los datos como 1 y 2 o como *varón* y *mujer*.

Supuestos. En algunos estadísticos y medidas se asume que hay unas categorías ordenadas (datos ordinales) o unos valores cuantitativos (datos de intervalos o de proporciones), como se explica en la sección sobre los estadísticos. Otros estadísticos son válidos cuando las variables de la tabla tienen categorías no ordenadas (datos nominales). Para los estadísticos basados en chi-cuadrado (phi, *V* de Cramér y coeficiente de contingencia), los datos deben ser una muestra aleatoria de una distribución multinomial.

Nota: las variables ordinales pueden ser códigos numéricos que representen categorías (por ejemplo, 1 = *bajo*, 2 = *medio*, 3 = *alto*) o valores de cadena. Sin embargo, se supone que el orden alfabético de los valores de cadena indica el orden correcto de las categorías. Por ejemplo, en una variable de cadena cuyos valores sean *bajo*, *medio*, *alto*, se interpreta el orden de las categorías como *alto*, *bajo*, *medio* (orden que no es el correcto). Por norma general, se puede indicar que es más fiable utilizar códigos numéricos para representar datos ordinales.

Para obtener tabulaciones cruzadas

1. En los menús seleccione:

   Analizar > **Estadísticos descriptivos** > **Tablas cruzadas...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables de fila y una o más variables de columna.

Si lo desea, puede:

- Seleccionar una o más variables de control.
- Pulsar **Estadísticos** para obtener pruebas y medidas de asociación para tablas o subtablas bidimensionales.
- Pulsar **Casillas** para obtener porcentajes, residuos y valores esperados y observados.
- Pulse **Formato** para controlar el orden de las categorías.

Capas de las tablas cruzadas

Si se seleccionan una o más variables de capas, se generará una tabulación cruzada por cada categoría de cada variable de capas (variable de control). Por ejemplo, si emplea una variable de fila, una variable de columna y una variable de capas con dos categorías, obtendrá una tabla bidimensional por cada categoría de la variable de capas. Para crear otra capa de variables de control, pulse en **Siguiente**. Se crean subtablas para cada combinación de categorías para cada variable de la 1ª capa, cada variable de la 2ª capa, y así sucesivamente. Si se solicitan estadísticos y medidas de asociación, se aplicarán sólo a las tablas bidimensionales.

Gráficos de barras agrupadas

Mostrar los gráficos de barras agrupadas. Los gráficos de barras agrupadas ayudan a resumir los datos por grupos de casos. Hay una agrupación de barras por cada valor de la variable especificada en el cuadro Filas. La variable que define las barras dentro de cada agrupación es la variable especificada en el cuadro Columnas. Por cada valor de esta variable hay un conjunto de barras de distinto color o trama. Si especifica más de una variable en Columnas o en Filas, se generará un gráfico de barras agrupadas por cada combinación de dos variables.

Tablas cruzadas mostrando variables de capa en capas de tabla

Mostrar variables de capa en capas de tabla. Puede seleccionar visualizar las variables de capa (variables de control) como capas en la tabla de tabulación cruzada. De esta forma podrá crear vistas que muestren los estadísticos globales de las variables de fila y columna y que permitan la obtención de detalles de las categorías de las variables de capa.

A continuación se muestra un ejemplo que utiliza el archivo de datos *demo.sav* (disponible en el directorio Samples del directorio de instalación) y que se ha obtenido de la siguiente forma:

1. Seleccione *Categoría de ingresos en miles (cating)* como la variable de fila, *Tiene PDA (pda)* como la variable de columna y *Nivel educativo (educ)* como la variable de capa.
2. Seleccione **Mostrar variables de capa en capas de tabla**.
3. Seleccione **Columna** en el cuadro de diálogo subordinado Mostrar en las casillas.
4. Ejecute el procedimiento de Tablas cruzadas, pulse dos veces en la tabla de tabulación cruzada y seleccione **Titulación universitaria** de la lista desplegable Nivel de estudios.

La vista seleccionada de la tabla de tabulación cruzada muestra los estadísticos de encuestados que tienen un título universitario.

Estadísticos de tablas cruzadas

Chi-cuadrado. Para las tablas con dos filas y dos columnas, seleccione **Chi-cuadrado** para calcular el chi-cuadrado de Pearson, el chi-cuadrado de la razón de verosimilitud, la prueba exacta de Fisher y el chi-cuadrado corregido de Yates (corrección por continuidad). Para tablas de 2 × 2, la prueba exacta de Fisher se calcula cuando una tabla que no resulta de filas o columnas que faltan en una tabla más grande tiene una celda con una frecuencia esperada de menos de 5. El chi-cuadrado corregido de Yates se calcula para todas las demás tablas de 2 × 2. Para las tablas con cualquier número de filas y columnas, seleccione **Chi-cuadrado** para calcular el chi-cuadrado de Pearson y el chi-cuadrado de la razón de verosimilitud. Cuando ambas variables de tabla son cuantitativas, **Chi-cuadrado** da como resultado la prueba de asociación lineal por lineal.

Correlaciones. Para las tablas en las que tanto las columnas como las filas contienen valores ordenados, **Correlaciones** da como resultado rho, el coeficiente de correlación de Spearman (sólo datos numéricos). La rho de Spearman es una medida de asociación entre órdenes de rangos. Cuando ambas variables de tabla (factores) son cuantitativas, **Correlaciones** da como resultado *r*, el coeficiente de correlación de Pearson, una medida de asociación lineal entre las variables.

Nominal. Para los datos nominales (sin orden intrínseco, como católico, protestante o judío), puede seleccionar el **Coeficiente de contingencia**, **Phi** (coeficiente) **y V de Cramér**, **Lambda** (lambdas simétricas y asimétricas y tau de Kruskal y Goodman) y el **Coeficiente de incertidumbre**.

- *Coeficiente de contingencia*.Medida de asociación basada en chi-cuadrado. El valor varía entre 0 y 1. El valor 0 indica que no hay asociación entre las variables de fila y de columna. Los valores cercanos a 1 indican que hay gran relación entre las variables. El valor máximo posible depende del número de filas y columnas de la tabla.
- *Phi y V de Cramer*. Phi es una medida de asociación basada en chi-cuadrado que implica dividir el estadístico de chi-cuadrado por el tamaño de la muestra y tomar la raíz cuadrada del resultado. V de Cramer es una medida de asociación basada en chi-cuadrado.

- *Lambda*. Medida de asociación que refleja la reducción proporcional del error cuando se utilizan valores de la variable independiente para predecir los valores de la variable dependiente. Un valor igual a 1 significa que la variable independiente pronostica perfectamente la variable dependiente. Un valor igual a 0 significa que la variable independiente no ayuda a pronosticar la variable dependiente.
- *Coeficiente de incertidumbre*. Medida de asociación que indica la reducción proporcional del error cuando se utilizan valores de una variable para predecir valores de la otra variable. Por ejemplo, un valor de 0,83 indica que el conocimiento de una variable reduce en un 83% el error al pronosticar los valores de la otra variable. El programa calcula tanto la versión simétrica como la asimétrica del coeficiente de incertidumbre.

Ordinal. Para las tablas en las que tanto las filas como las columnas contienen valores ordenados, seleccione **Gamma** (orden cero para tablas bidimensionales y condicional para tablas cuyo factor de clasificación va de 3 a 10), **Tau-b de Kendall** y **Tau-c de Kendall**. Para pronosticar las categorías de columna de las categorías de fila, seleccione **d de Somers**.

- *Gamma*. Una medida simétrica de asociación entre dos variables ordinales con los rangos de -1 y 1. Los valores que se aproximan a un valor absoluto de 1 indican una relación fuerte entre las dos variables. Los valores próximos a cero indican que hay poca o ninguna relación entre las dos variables. Para las tablas bidimensionales, se muestran las gammas de orden cero. Para las tablas de tres o más factores de clasificación, se muestran las gammas condicionales.
- *d de Somers*. Una medida de asociación entre dos variables ordinales en el rango de -1 a 1. Los valores cercanos a un valor absoluto de 1 indican una relación fuerte entre las dos variables, y los valores cercanos a 0 indican poca o ninguna relación entre las variables. La d de Somers es una extensión asimétrica de gamma que difiere sólo en la inclusión del número de pares no empatados en la variable independiente. También se calcula una versión no simétrica de este estadístico.
- *Tau-b de Kendall*. Medida no paramétrica de correlación para variables ordinales o clasificadas que tienen en cuenta los empates. El signo del coeficiente indica la dirección de la relación y su valor absoluto indica la fuerza de la relación. Los valores mayores indican que la relación es más estrecha. Los valores posibles van de -1 a 1, pero un valor de -1 o +1 sólo se puede obtener a partir de tablas cuadradas.
- *Tau-c de Kendall*. Una medida no paramétrica de asociación para variables ordinales que ignora los empates. El signo del coeficiente indica la dirección de la relación y su valor absoluto indica la fuerza de la relación. Los valores mayores indican que la relación es más estrecha. Los valores posibles van de -1 a 1, pero un valor de -1 o +1 sólo se puede obtener a partir de tablas cuadradas.

Nominal por intervalo. Cuando una variable es categórica y la otra es cuantitativa, seleccione **Eta**. La variable categórica debe codificarse numéricamente.

- *Eta*. Medida de asociación que va de 0 a 1, con 0 que indica que no hay asociación entre las variables de fila y columna y valores próximos a 1 que indican un alto grado de asociación. Eta resulta apropiada para una variable dependiente medida en una escala de intervalo (por ejemplo, ingresos) y una variable independiente con un número limitado de categorías (por ejemplo, género). Se calculan dos valores eta: uno trata la variable de las filas como una variable de intervalo; el otro trata la variable de las columnas como una variable de intervalo.

Kappa. La kappa de Cohen mide el acuerdo entre las evaluaciones de dos evaluadores cuando ambos califican el mismo objeto. Un valor igual a 1 indica un acuerdo perfecto. Un valor igual a 0 indica que el acuerdo no es mejor que el que se obtendría por azar. Kappa se basa en una tabla cuadrada en la que los valores de filas y columnas representan la misma escala. Cualquier celda que tenga valores observado para una variable pero no para la otra tiene asignado un recuento de 0. Kappa no se calcula si el tipo de almacenamiento de datos (serie o numérico) no es el mismo para las dos variables. Para una variable de cadena, ambas variables deben tener la misma longitud definida.

Riesgo. Para tablas 2 x 2, una medida de la fuerza de la asociación entre la presencia de un factor y la aparición de un suceso. Si el intervalo de confianza para el estadístico incluye un valor de 1, no se podrá asumir que el factor está asociado con el evento. Cuando la ocurrencia del factor es poco común, se puede utilizar la razón de probabilidad como estimación o riesgo relativo.

McNemar. Una prueba no paramétrica para dos variables dicotómicas relacionadas. Contrasta los cambios de respuesta utilizando una distribución chi-cuadrado. Es útil para detectar cambios en las

respuestas causadas por la intervención experimental en los diseños del tipo "antes-después". Para las tablas cuadradas de mayor orden se informa de la prueba de simetría de McNemar-Bowker.

Estadísticos de Cochran y Mantel-Haenszel. Cochran y Mantel-Haenszel se pueden utilizar para contrastar la independencia entre una variable de factor dicotómica y una variable de respuesta dicotómica, condicionada a patrones de covariables definidos por una o más variables de capa (control). Tenga en cuenta que mientras que otros estadísticos se calculan capa por capa, los estadísticos de Cochran y Mantel-Haenszel se calculan una sola vez para todas las capas.

Visualización en casillas de tablas cruzadas

Para ayudarle a descubrir las tramas en los datos que contribuyen a una prueba de chi-cuadrado significativa, el procedimiento Tablas cruzadas muestra las frecuencias esperadas y tres tipos de residuos (desviaciones) que miden la diferencia entre las frecuencias observadas y las esperadas. Cada casilla de la tabla puede contener cualquier combinación de recuentos, porcentajes y residuos seleccionados.

Recuentos. El número de casos realmente observados y el número de casos esperados si las variables de fila y columna son independientes entre sí. Puede optar por ocultar recuentos inferiores un número entero especificado. Los valores ocultos se visualizarán como **<N**, donde **N** es el entero especificado. El entero especificado debe ser mayor o igual a 2, aunque se permite el valor 0 y especifica que no se hay recuentos ocultos.

Comparar las proporciones de columna. Esta opción calcula comparaciones por pares de proporciones de columnas e indica los pares de columnas (de una fila concreta) que son significativamente diferentes. Las diferencias significativas se indican en la tabla cruzada con formato de estilo APA utilizando subíndices de letras y se calculan con un nivel de significación de 0,05. *Nota*: si se especifica esta opción sin seleccionar recuentos observados o porcentajes de columnas, se incluirán los recuentos observados en la tabla cruzada, con subíndices de estilo APA indicando los resultados de las pruebas de proporciones de columnas.

- **Corregir valores p (método de Bonferroni).** Las comparaciones por parejas de las proporciones de columnas utilizan la corrección de Bonferroni, que ajusta el nivel de significación observado por el hecho de que se realizan múltiples comparaciones.

Porcentajes. Los porcentajes se pueden sumar a través de las filas o a lo largo de las columnas. También se encuentran disponibles los porcentajes del número total de casos representados en la tabla (una capa).

Nota: Si **Ocultar recuentos pequeños** está seleccionado en el grupo Recuentos, se ocultarán también los porcentajes asociados con recuentos ocultos.

Residuos. Los residuos brutos no tipificados presentan la diferencia entre los valores observados y los esperados. También se encuentran disponibles los residuos estandarizados y estandarizados corregidos.

- *Sin estandarizar*. La diferencia entre un valor observado y el valor esperado. El valor esperado es el número de casos que se esperaría encontrar en la casilla si no hubiera relación entre las dos variables. Un residuo positivo indica que hay más casos en la casilla de los que habría en ella si las variables de fila y columna fueran independientes.
- *Estandarizado*. El residuo dividido por una estimación de su desviación estándar. Los residuos estandarizados, que son conocidos también como los residuos de Pearson o residuos estandarizados, tienen una media de 0 y una desviación estándar de 1.
- *Estandarizado ajustado*. El residuo de una casilla (el valor observado menos el valor esperado) dividido por una estimación de su error estándar. El residuo estandarizado resultante viene expresado en unidades de desviación estándar, por encima o por debajo de la media.

Crear tabla de estilo APA. Crea tablas de salida que cumplen las directrices de estilo APA.

Nota: Las opciones **Observado**, **Esperado**, **Fila**, **Columna**y **Total** no están disponibles cuando se selecciona **Crear tabla de estilo APA**.

Ponderaciones no enteras. Los recuentos de las casillas suelen ser valores enteros, ya que representan el número de casos de cada casilla. Sin embargo, si el archivo de datos está ponderado en un momento

determinado por una variable de ponderación con valores fraccionarios (por ejemplo, 1,25), los recuentos de las casillas pueden que también sean valores fraccionarios. Puede truncar o redondear estos valores antes o después de calcular los recuentos de las casillas o bien utilizar recuentos de casillas fraccionarios en la presentación de las tablas y los cálculos de los estadísticos.

- *Redondear recuentos de casillas*. Las ponderaciones de los casos se utilizan tal cual, pero las ponderaciones acumuladas en las casillas se redondean antes de calcular cualquiera de los estadísticos.
- *Truncar recuentos de casillas*. Las ponderaciones de los casos se utilizan tal cual, pero las ponderaciones acumuladas en las casillas se truncan antes de calcular cualquiera de los estadísticos.
- *Redondear ponderaciones de casos*. Se redondean las ponderaciones de los casos antes de utilizarlas.
- *Truncar ponderaciones de casos*. Se truncan las ponderaciones de los casos antes de utilizarlas.
- *Sin ajustes*. Las ponderaciones de casos se utilizan tal cual y se utilizan recuentos de casillas fraccionadas. Sin embargo, cuando se solicitan Estadísticos exactos (disponibles sólo con Muestreo y Pruebas), las ponderaciones acumuladas en las casillas se truncan o redondean antes de calcular los estadísticos de las Pruebas exactas.

Formato de tablas cruzadas

Puede ordenar las filas en orden ascendente o descendente de los valores de la variable de fila.

Resumir

El procedimiento Resumir calcula estadísticos de subgrupo para las variables dentro de las categorías de una o más variables de agrupación. Se cruzan todos los niveles de las variables de agrupación. Puede elegir el orden en el que se mostrarán los estadísticos. También se muestran estadísticos de resumen para cada variable a través de todas las categorías. Los valores de los datos en cada categoría pueden mostrarse en una lista o suprimirse. Con grandes conjuntos de datos, tiene la opción de listar sólo los primeros *n* casos.

Ejemplo. ¿Cuál es la media de las ventas por regiones o por tipo de cliente? Podrá descubrir que el importe medio de las ventas es ligeramente superior en la región occidental respecto a las demás regiones, y que la media más alta se da entre los clientes de empresas privadas de la zona occidental .

Estadísticas. Suma, número de casos, media, mediana, mediana agrupada, error estándar de la media, mínimo, máximo, rango, valor de la variable para la primera categoría de la variable de agrupación, valor de la variable para la última categoría de la variable de agrupación, desviación estándar, varianza, curtosis, error estándar de curtosis, asimetría, error estándar de asimetría, porcentaje de la suma total, porcentaje del *N* total, porcentaje de la suma en, porcentaje de *N* en, media geométrica y media armónica.

Resumir: Consideraciones sobre los datos

Datos. Las variables de agrupación son variables categóricas cuyos valores pueden ser numéricos o de cadena. El número de categorías debe ser razonablemente pequeño. Las otras variables deben poder ordenarse mediante rangos.

Supuestos. Algunos de los estadísticos opcionales de subgrupo, como la media y la desviación estándar, se basan en la teoría normal y son adecuados para variables cuantitativas con distribuciones simétricas. Los estadísticos robustos, tales como la mediana y el rango, son adecuados para las variables cuantitativas que pueden o no cumplir el supuesto de normalidad.

Para obtener resúmenes de casos

1. En los menús seleccione:

   Analizar > **Informes** > **Resúmenes de casos...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables.

Si lo desea, puede:

- Seleccionar una o más variables de agrupación para dividir los datos en subgrupos.
- Pulsar **Opciones** para cambiar el título de los resultados, añadir un texto al pie debajo de los resultados o excluir los casos con valores perdidos.
- Pulse **Estadísticas** para obtener estadísticas opcionales.
- Seleccionar **Mostrar los casos** para listar los casos en cada subgrupo. De forma predeterminada, el sistema enumera sólo los 100 primeros casos del archivo. Puede aumentar o disminuir el valor de **Limitar los casos a los primeros** *n* o desactivar ese elemento para enumerar todos los casos.

Resumir: Opciones

Resumir permite cambiar el título de los resultados o añadir un texto que aparecerá debajo de la tabla de resultados. Puede controlar el ajuste de las líneas en los títulos y textos escribiendo \n en el lugar donde desee insertar una línea de separación.

Además, puede elegir entre mostrar o suprimir los subtítulos para los totales e incluir o excluir los casos con valores perdidos para cualquiera de las variables utilizadas en cualquiera de los análisis. A menudo es aconsejable representar los casos perdidos en los resultados con un punto o un asterisco. Introduzca un carácter, frase o código que desee que aparezca cuando haya un valor perdido; de lo contrario, no se aplicará ningún tratamiento especial a los casos perdidos en los resultados.

Resumir: Estadísticos

Puede elegir uno o más de los siguientes estadísticos de subgrupo para las variables dentro de cada categoría de cada variable de agrupación: suma, número de casos, media, mediana, mediana agrupada, error estándar de la media, mínimo, máximo, rango, valor de la variable para la primera categoría de la variable de agrupación, valor de la variable para la última categoría de la variable de agrupación, desviación estándar, varianza, curtosis, error estándar de curtosis, asimetría, error estándar de asimetría, porcentaje de la suma total, porcentaje del *N* total, porcentaje de la suma en, porcentaje de *N* en, media geométrica y media armónica. El orden en el que aparecen los estadísticos en la lista Estadísticos de casilla es el orden en el que se mostrarán en los resultados. También se muestran estadísticos de resumen para cada variable a través de todas las categorías.

Primero. Muestra el primer valor de los datos encontrado en el archivo de datos.

Media geométrica. La raíz enésima del producto de los valores de los datos, donde n representa el número de casos.

Mediana agrupada. Mediana que se calcula para los datos codificados en grupos. Por ejemplo, con datos de edades, si cada valor de los 30 se ha codificado como 35, cada valor de los 40 como 45 y así sucesivamente, la mediana agrupada es la mediana calculada a partir de los datos codificados.

media armónica. Se utiliza para estimar un tamaño de grupo promedio cuando los tamaños de muestra de los grupos no son iguales. La media armónica es el número total de muestras divido por la suma de los inversos de los tamaños de las muestras.

Curtosis. Medida del grado en que hay valores atípicos. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que los datos muestran más valores atípico extremos que una distribución normal. La curtosis negativa indica que los datos presentan valores atípicos menos extremados que una distribución normal. Algunas veces se hace referencia a la definición de curtosis que se utiliza, donde el valor es 0 para una distribución normal, como curtosis de exceso. Algunos programas pueden notificar curtosis de tal manera que el valor es 3 para una distribución normal.

Último. Muestra el último valor de los datos encontrado en el archivo de datos.

máximo. El valor mayor de una variable numérica.

Media. Medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.

Mediana. El valor por encima y por debajo del cual se sitúan la mitad de los casos: el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).

Mínimo. El valor menor de una variable numérica.

N. Número de casos (observaciones o registros).

Porcentaje del N total. Porcentaje del número total de casos en cada categoría.

Porcentaje de suma total. Porcentaje de la suma total en cada categoría.

Rango. Diferencia entre los valores mayor y menor de una variable numérica; el máximo menos el mínimo.

Asimetría. Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de sesgo de 0. Una distribución con un sesgo positivo significativo tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error estándar se asume que indica una desviación de la simetría.

Desviación estándar. Una medida de dispersión sobre la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación estándar de la media y el 95% queda entre dos desviaciones estándar. Por ejemplo, si la edad media es de 45 años, con una desviación estándar de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.

Error estándar de curtosis. La relación entre la curtosis y su error estándar puede utilizarse como prueba de normalidad (es decir, se puede rechazar la normalidad si la relación es inferior a -2 o superior a +2). Un valor grande y positivo para la curtosis indica que las colas son más largas que las de una distribución normal; por el contrario, un valor extremo y negativo indica que las colas son más cortas (llegando a tener forma de caja como en la distribución uniforme).

Error estándar de la media. Medida de cuánto puede variar el valor de la media de una muestra a otra tomada de la misma distribución. Puede utilizarse para comparar de forma aproximada la media observada respecto a un valor hipotetizado (es decir, se puede concluir que los dos valores son distintos si la diferencia entre ellos, dividida por el error estándar, es menor que -2 o mayor que +2).

Error estándar de asimetría. La relación entre la asimetría y su error estándar puede utilizarse como prueba de normalidad (es decir, se puede rechazar la normalidad si la relación es inferior a -2 o superior a +2). Un valor grande y positivo para la asimetría indica una cola larga a la derecha; un valor extremo y negativo indica una cola larga por la izquierda

Suma. Suma o total de los valores, a lo largo de todos los casos con valores no perdidos.

Varianza. Medida de dispersión sobre la media, igual a la suma de las desviaciones al cuadrado de la media divida por el número de casos menos uno. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.

Medias

El procedimiento Medias calcula medias de subgrupo y estadísticos univariados relacionados para variables dependientes dentro de las categorías de una o más variables independientes. Si lo desea, puede obtener el análisis de varianza de un factor, la eta y pruebas de linealidad.

Ejemplo. Mida la cantidad media de grasa absorbida en función de tres tipos distintos de aceite comestible y realice un análisis de varianza de un factor para comprobar si difieren las medias.

Estadísticas. Suma, número de casos, media, mediana, mediana agrupada, error estándar de la media, mínimo, máximo, rango, valor de la variable para la primera categoría de la variable de agrupación, valor de la variable para la última categoría de la variable de agrupación, desviación estándar, varianza, curtosis, error estándar de curtosis, asimetría, error estándar de asimetría, porcentaje de la suma total, porcentaje del *N* total, porcentaje de la suma en, porcentaje de *N* en, media geométrica y media

armónica. Las opciones incluyen: análisis de varianza, eta, eta cuadrado y pruebas de linealidad de *R* y R^2.

Medias: Consideraciones sobre los datos

Datos. Las variables dependientes son cuantitativas y las independientes son categóricas. Los valores de las variables categóricas pueden ser numéricos o de cadena.

Supuestos. Algunos de los estadísticos opcionales de subgrupo, como la media y la desviación estándar, se basan en la teoría normal y son adecuados para variables cuantitativas con distribuciones simétricas. Los estadísticos robustos, tales como la mediana son adecuados para las variables cuantitativas que pueden o no cumplir el supuesto de normalidad. El análisis de varianza es robusto a las desviaciones de la normalidad, aunque los datos de cada casilla deberían ser simétricos. El análisis de varianza también supone que los grupos proceden de poblaciones con la misma varianza. Para comprobar este supuesto, utilice la prueba de homogeneidad de las varianzas de Levene, disponible en el procedimiento ANOVA de un factor.

Para obtener medias de subgrupo

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Medias...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables dependientes.
3. Utilice uno de los siguientes métodos para seleccionar variables independientes categóricas:
   - Seleccionar una o más variables independientes. Se mostrarán resultados individuales para cada variable independiente.
   - Seleccione una o más capas de variables independientes. Cada capa subdivide consecutivamente la muestra. Si tiene una variable independiente en Capa 1 y otra variable independiente en Capa 2, los resultados se mostrarán en una tabla cruzada en contraposición a tablas individuales para cada variable independiente.
4. Si lo desea, pulse en **Opciones** si desea obtener estadísticos opcionales, una tabla de análisis de varianza, eta, eta cuadrado, *R*, y R^2.

Medias: Opciones

Puede elegir uno o más de los siguientes estadísticos de subgrupo para las variables dentro de cada categoría de cada variable de agrupación: suma, número de casos, media, mediana, mediana agrupada, error estándar de la media, mínimo, máximo, rango, valor de la variable para la primera categoría de la variable de agrupación, valor de la variable para la última categoría de la variable de agrupación, desviación estándar, varianza, curtosis, error estándar de curtosis, asimetría, error estándar de asimetría, porcentaje de la suma total, porcentaje del *N* total, porcentaje de la suma en, porcentaje de *N* en, media geométrica, media armónica. Se puede cambiar el orden de aparición de los estadísticos de subgrupo. El orden en el que aparecen en la lista Estadísticos de casilla es el mismo orden que presentarán en el resultado. También se muestran estadísticos de resumen para cada variable a través de todas las categorías.

Primero. Muestra el primer valor de los datos encontrado en el archivo de datos.

Media geométrica. La raíz enésima del producto de los valores de los datos, donde n representa el número de casos.

Mediana agrupada. Mediana que se calcula para los datos codificados en grupos. Por ejemplo, con datos de edades, si cada valor de los 30 se ha codificado como 35, cada valor de los 40 como 45 y así sucesivamente, la mediana agrupada es la mediana calculada a partir de los datos codificados.

media armónica. Se utiliza para estimar un tamaño de grupo promedio cuando los tamaños de muestra de los grupos no son iguales. La media armónica es el número total de muestras divido por la suma de los inversos de los tamaños de las muestras.

Curtosis. Medida del grado en que hay valores atípicos. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que los datos muestran más valores atípico extremos que una distribución normal. La curtosis negativa indica que los datos presentan valores atípicos menos extremados que una distribución normal. Algunas veces se hace referencia a la definición de curtosis que se utiliza, donde el valor es 0 para una distribución normal, como curtosis de exceso. Algunos programas pueden notificar curtosis de tal manera que el valor es 3 para una distribución normal.

Último. Muestra el último valor de los datos encontrado en el archivo de datos.

máximo. El valor mayor de una variable numérica.

Media. Medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.

Mediana. El valor por encima y por debajo del cual se sitúan la mitad de los casos: el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).

Mínimo. El valor menor de una variable numérica.

N. Número de casos (observaciones o registros).

Porcentaje del N total. Porcentaje del número total de casos en cada categoría.

Porcentaje de la suma total. Porcentaje de la suma total en cada categoría.

Rango. Diferencia entre los valores mayor y menor de una variable numérica; el máximo menos el mínimo.

Asimetría. Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de sesgo de 0. Una distribución con un sesgo positivo significativo tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error estándar se asume que indica una desviación de la simetría.

Desviación estándar. Una medida de dispersión sobre la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación estándar de la media y el 95% queda entre dos desviaciones estándar. Por ejemplo, si la edad media es de 45 años, con una desviación estándar de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.

Error estándar de curtosis. La relación entre la curtosis y su error estándar puede utilizarse como prueba de normalidad (es decir, se puede rechazar la normalidad si la relación es inferior a -2 o superior a +2). Un valor grande y positivo para la curtosis indica que las colas son más largas que las de una distribución normal; por el contrario, un valor extremo y negativo indica que las colas son más cortas (llegando a tener forma de caja como en la distribución uniforme).

Error estándar de la media. Medida de cuánto puede variar el valor de la media de una muestra a otra tomada de la misma distribución. Puede utilizarse para comparar de forma aproximada la media observada respecto a un valor hipotetizado (es decir, se puede concluir que los dos valores son distintos si la diferencia entre ellos, dividida por el error estándar, es menor que -2 o mayor que +2).

Error estándar de asimetría. La relación entre la asimetría y su error estándar puede utilizarse como prueba de normalidad (es decir, se puede rechazar la normalidad si la relación es inferior a -2 o superior a +2). Un valor grande y positivo para la asimetría indica una cola larga a la derecha; un valor extremo y negativo indica una cola larga por la izquierda

Suma. Suma o total de los valores, a lo largo de todos los casos con valores no perdidos.

Varianza. Medida de dispersión sobre la media, igual a la suma de las desviaciones al cuadrado de la media divida por el número de casos menos uno. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.

Estadísticos para la primera capa

Eta y tabla Anova. Muestra una tabla de análisis de varianza de un factor y calcula la eta y la eta cuadrado (medidas de asociación) para cada variable independiente de la primera capa.

Prueba de linealidad. Calcula la suma de cuadrados, grados de libertad y la media cuadrática asociada con componentes lineales y no lineales, así como la razón F, R y R cuadrado. Si la variable independiente es una cadena corta entonces la linealidad no se calcula.

Cubos OLAP

El procedimiento Cubos OLAP (siglas del inglés On-Line Analytic Processing, «Procesamiento analítico interactivo») calcula totales, medias y otros estadísticos univariantes para variables de resumen continuas dentro de las categorías de una o más variables categóricas de agrupación. En la tabla se creará una nueva capa para cada categoría de cada variable de agrupación.

Ejemplo. El total y el promedio de ventas para diversas regiones y líneas de producto, dentro de las regiones.

Estadísticas. Suma, número de casos, media, mediana, mediana agrupada, error estándar de la media, mínimo, máximo, rango, valor de la variable para la primera categoría de la variable de agrupación, valor de la variable para la última categoría de la variable de agrupación, desviación estándar, varianza, curtosis, error estándar de curtosis, asimetría, error estándar de asimetría, porcentaje de casos totales, porcentaje de la suma total, porcentaje de casos totales dentro de las variables agrupadas, porcentaje de la suma total dentro de las variables agrupadas, media geométrica y media armónica.

Cubos OLAP: Consideraciones sobre los datos

Datos. Las variables de resumen son cuantitativas (variables continuas medidas en una escala de intervalo o de razón) y las variables de agrupación son categóricas. Los valores de las variables categóricas pueden ser numéricos o de cadena.

Supuestos. Algunos de los estadísticos opcionales de subgrupo, como la media y la desviación estándar, se basan en la teoría normal y son adecuados para variables cuantitativas con distribuciones simétricas. Los estadísticos robustos, tales como la mediana y el rango, son adecuados para las variables cuantitativas que pueden o no cumplir el supuesto de normalidad.

Para obtener cubos OLAP

1. En los menús seleccione:

   Analizar > **Informes** > **Cubos OLAP...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables de resumen continuas.
3. Seleccione una o más variables de agrupación categóricas.

Si lo desea:

- Seleccionar diferentes estadísticos de resumen (pulse **Estadísticos**). Debe seleccionar una o más variables de agrupación para poder seleccionar estadísticos de resumen.
- Calcule las diferencias existentes entre los pares de variables y los pares de grupos definidos por una variable de agrupación (pulse **Diferencias**).
- Crear títulos de tabla personalizados (pulse **Título**).
- Oculta recuentos que sean inferiores a un entero especificado. Los valores ocultos se visualizarán como **<N**, donde **N** es el entero especificado. El número entero especificado debe ser mayor o igual a 2.

Cubos OLAP: Estadísticos

Puede elegir uno o varios de los siguientes estadísticos de subgrupo para las variables de resumen dentro de cada categoría de cada variable de agrupación: Suma, Número de casos, Media, Mediana, Mediana agrupada, Error estándar de la media, Mínimo, Máximo, Rango, Valor de la variable para la primera categoría de la variable de agrupación, Valor de la variable para la última categoría de la variable de agrupación, Desviación estándar, Varianza, Curtosis, Error estándar de curtosis, Asimetría, Error estándar de asimetría, Porcentaje de casos totales, Porcentaje de la suma total, Porcentaje de casos totales dentro

de las variables de agrupación, Porcentaje de la suma total dentro de las variables de agrupación, Media geométrica y Media armónica.

Se puede cambiar el orden de aparición de los estadísticos de subgrupo. El orden en el que aparecen en la lista Estadísticos de casilla es el mismo orden que presentarán en el resultado. También se muestran estadísticos de resumen para cada variable a través de todas las categorías.

Primero. Muestra el primer valor de los datos encontrado en el archivo de datos.

Media geométrica. La raíz enésima del producto de los valores de los datos, donde n representa el número de casos.

Mediana agrupada. Mediana que se calcula para los datos codificados en grupos. Por ejemplo, con datos de edades, si cada valor de los 30 se ha codificado como 35, cada valor de los 40 como 45 y así sucesivamente, la mediana agrupada es la mediana calculada a partir de los datos codificados.

media armónica. Se utiliza para estimar un tamaño de grupo promedio cuando los tamaños de muestra de los grupos no son iguales. La media armónica es el número total de muestras divido por la suma de los inversos de los tamaños de las muestras.

Curtosis. Medida del grado en que hay valores atípicos. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que los datos muestran más valores atípico extremos que una distribución normal. La curtosis negativa indica que los datos presentan valores atípicos menos extremados que una distribución normal. Algunas veces se hace referencia a la definición de curtosis que se utiliza, donde el valor es 0 para una distribución normal, como curtosis de exceso. Algunos programas pueden notificar curtosis de tal manera que el valor es 3 para una distribución normal.

Último. Muestra el último valor de los datos encontrado en el archivo de datos.

máximo. El valor mayor de una variable numérica.

Media. Medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.

Mediana. El valor por encima y por debajo del cual se sitúan la mitad de los casos: el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).

Mínimo. El valor menor de una variable numérica.

N. Número de casos (observaciones o registros).

Porcentaje del N en. Porcentaje del número de casos para la variable de agrupación especificada dentro de las categorías de otras variables de agrupación. Si sólo tiene una variable de agrupación, este valor es idéntico al porcentaje del número de casos total.

Porcentaje de suma en. Porcentaje de la suma para la variable de agrupación especificada dentro de las categorías de otras variables de agrupación. Si sólo tiene una variable de agrupación, este valor es idéntico al porcentaje de la suma total.

Porcentaje del N total. Porcentaje del número total de casos en cada categoría.

Porcentaje de suma total. Porcentaje de la suma total en cada categoría.

Rango. Diferencia entre los valores mayor y menor de una variable numérica; el máximo menos el mínimo.

Asimetría. Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de sesgo de 0. Una distribución con un sesgo positivo significativo tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error estándar se asume que indica una desviación de la simetría.

Desviación estándar. Una medida de dispersión sobre la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación estándar de la media y el 95% queda entre dos

desviaciones estándar. Por ejemplo, si la edad media es de 45 años, con una desviación estándar de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.

Error estándar de curtosis. La relación entre la curtosis y su error estándar puede utilizarse como prueba de normalidad (es decir, se puede rechazar la normalidad si la relación es inferior a -2 o superior a +2). Un valor grande y positivo para la curtosis indica que las colas son más largas que las de una distribución normal; por el contrario, un valor extremo y negativo indica que las colas son más cortas (llegando a tener forma de caja como en la distribución uniforme).

Error estándar de la media. Medida de cuánto puede variar el valor de la media de una muestra a otra tomada de la misma distribución. Puede utilizarse para comparar de forma aproximada la media observada respecto a un valor hipotetizado (es decir, se puede concluir que los dos valores son distintos si la diferencia entre ellos, dividida por el error estándar, es menor que -2 o mayor que +2).

Error estándar de asimetría. La relación entre la asimetría y su error estándar puede utilizarse como prueba de normalidad (es decir, se puede rechazar la normalidad si la relación es inferior a -2 o superior a +2). Un valor grande y positivo para la asimetría indica una cola larga a la derecha; un valor extremo y negativo indica una cola larga por la izquierda

Suma. Suma o total de los valores, a lo largo de todos los casos con valores no perdidos.

Varianza. Medida de dispersión sobre la media, igual a la suma de las desviaciones al cuadrado de la media divida por el número de casos menos uno. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.

Cubos OLAP: Diferencias

Este cuadro de diálogo le permite calcular el porcentaje y las diferencias aritméticas entre las variables de resumen o entre los grupos definidos por una variable de agrupación. Las diferencias se calculan para todas las medidas seleccionadas en el cuadro de diálogo Cubos OLAP: Estadísticos.

Diferencias entre variables. Calcula las diferencias entre pares de variables. Los valores de los estadísticos de resumen para la segunda variable de cada par (la variable Menos) se restan de los valores de los estadísticos de resumen correspondientes a la primera variable del par. En cuanto a las diferencias porcentuales, el valor de la variable de resumen para la variable Menos es el que se usa como denominador. Debe seleccionar al menos dos variables de resumen en el cuadro de diálogo principal para poder especificar las diferencias entre las variables.

Diferencias entre grupos de casos. Calcula las diferencias entre pares de grupos definidos por una variable de agrupación. Los valores de los estadísticos de resumen para la segunda categoría de cada par (la variable Menos) se restan de los valores de los estadísticos de resumen correspondientes a la primera categoría del par. Las diferencias porcentuales utilizan el valor del estadístico de resumen de la categoría Menos como denominador. Debe seleccionar una o más variables de agrupación en el cuadro de diálogo principal para poder especificar las diferencias entre los grupos.

Cubos OLAP: Título

Puede cambiar el título de los resultados o añadir un texto al pie que aparecerá debajo de la tabla de resultados. También puede controlar el ajuste de las líneas de los títulos y de los textos al pie escribiendo \n en el lugar del texto donde desee insertar una línea de separación.

Proporciones

Introducción a las proporciones

Las proporciones calculan pruebas e intervalos de confianza para proporciones binomiales o diferencias de proporciones. Hay estadísticos disponibles para proporciones de una muestra (probadas frente a un valor especificado), muestras emparejadas (distintas variables) o muestras independientes (distintos grupos de casos). Se ofrecen diversas opciones para tipos de estadísticos de prueba e intervalos de confianza. Otros procedimientos que proporcionan algunas de las mismas funciones son **CROSSTABS**, **NPAR TESTS** y **NPTESTS**.

Proporciones de una muestra
Pruebas e intervalos de confianza de una muestra. La salida incluye la proporción observada, la estimación de la diferencia entre la proporción de población y la proporción de población de la hipótesis, errores estándar asintóticos bajo hipótesis nulas y alternativas, estadísticos de prueba especificados con probabilidades bilaterales, e intervalos de confianza especificados para las proporciones.

Proporciones de muestras emparejadas
Pruebas e intervalos de confianza de muestras emparejadas para las diferencias entre proporciones. La salida incluye las proporciones observadas, estimaciones de las diferencias en las proporciones de población, errores estándar asintóticos de las diferencias de población bajo hipótesis nulas y alternativas, estadísticos de prueba especificados con probabilidades bilaterales, e intervalos de confianza especificados para las diferencias en las proporciones.

Proporciones de muestras independientes
Pruebas e intervalos de confianza de muestras independientes. La salida incluye las proporciones observadas, estimaciones de las diferencias en las proporciones de población, errores estándar asintóticos de las diferencias de población bajo hipótesis nulas y alternativas, estadísticos de prueba especificados con probabilidades bilaterales, e intervalos de confianza especificados para las diferencias en las proporciones.

Proporciones de una muestra

El procedimiento Proporciones de una muestra proporciona pruebas e intervalos de confianza para proporciones binomiales individuales. Se presupone que los datos proceden de una única muestra aleatoria, y cada prueba de hipótesis o intervalo de confianza es una prueba independiente o un intervalo individual, basado en la proporción binomial. La salida incluye la proporción observada, la estimación de la diferencia entre la proporción de población y la proporción de población de la hipótesis, errores estándar asintóticos bajo hipótesis nulas y alternativas, estadísticos de prueba especificados con probabilidades bilaterales, e intervalos de confianza especificados para las proporciones.

Ejemplo

Estadísticos
Agresti-Coull, Anscombe, Clopper-Pearson (Exacto), Jeffreys, Logit, Wald, Wald (continuidad corregida), Puntuación de Wilson, Puntuación de Wilson (continuidad corregida), Binomial exacto, Binomial ajustado mid-p, Puntuación, Puntuación (continuidad corregida).

Consideraciones sobre los datos

Datos
El procedimiento muestra estadísticos de prueba solicitados y probabilidades bilaterales, intervalos de confianza para diferencias de proporciones, así como proporciones, errores estándar y recuentos para cada grupo o variable. El procedimiento está limitado a un máximo de un valor de prueba.

Supuestos

Obtención de pruebas de proporciones de una muestra

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de una muestra...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables de contraste cuantitativas.
3. Si lo desea, puede:
   - Seleccione valores de criterios de éxito en la sección **Definir éxito**:

Último valor
Se utiliza el valor más alto o último entre los valores distintos ordenados de los datos. Esto se aplica a variables numéricas o de cadena. Ésta es el ajuste predeterminado.

Primer valor
Se utiliza el valor más bajo o primero entre los valores distintos ordenados de los datos. Esto se aplica a variables numéricas o de cadena.

Valores
Uno o más valores específicos entre paréntesis. Si hay varios valores, se deben separar por espacios. Esto se aplica a variables numéricas o de cadena. Los valores de variable de cadena se deben delimitar entre comillas simples.

Punto medio
Los valores que están en la mitad o por encima del rango de valores observados de los datos. Esto se aplica únicamente a datos numéricos.

Punto de corte
Valores que están en un valor especificado o por encima de él. Esto se aplica únicamente a datos numéricos.

- Haga clic en **Intervalos de confianza...** para especificar qué tipos de intervalos de confianza se muestran, o para suprimir todos los intervalos de confianza.
- Haga clic en **Pruebas...** para especificar qué tipos de estadísticas de pruebas se muestran, o para suprimir todas las pruebas.
- Haga clic en **Valores perdidos...** para controlar el tratamiento de los datos que faltan.
- Haga clic en **Bootstrap...** para derivar estimaciones robustas de errores estándar e intervalos de confianza para estimaciones como la media, la mediana, la proporción, la odds ratio, el coeficiente de correlación o el coeficiente de regresión.

4. Pulse **Aceptar**.

Proporciones de una muestra: Intervalos de confianza

El diálogo Intervalos de confianza proporciona opciones para especificar el nivel de cobertura y para seleccionar qué tipos de intervalos de confianza se muestran.

Nivel de cobertura
Especifica el porcentaje del intervalo de confianza. Se debe especificar un valor numérico en el rango (0,100). 95 es el valor predeterminado.

Tipos de intervalos
Proporciona opciones para especificar qué tipos de intervalos de confianza se muestran. Las opciones disponibles son:

- Agresti-Coull
- Anscombe
- Clopper-Pearson (exacto)
- Jeffreys
- Logit
- Wald
- Wald (Continuidad corregida)
- Puntuación Wilson
- Puntuación Wilson (Continuidad corregida)

Especificación de intervalos de confianza para proporciones de una muestra

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de una muestra...**

2. Pulse **Intervalos de confianza** para especificar qué tipos de intervalos de confianza se muestran, o para suprimir todos los intervalos de confianza.

Proporciones de una muestra: Pruebas

El diálogo Pruebas proporciona opciones para especificar qué tipos de estadísticos de prueba se muestran.

Todos
Se muestran todas las estadísticos de prueba en los resultados.

Ninguna
No se muestran estadísticos de prueba en los resultados.

Binomial exacto
Muestra probabilidades binomiales exactas.

Binomial ajustado mid-p
Muestra probabilidades binomiales ajustadas mid-p. Este es un valor predeterminado.

Puntuación
Muestra la estadística de prueba Z de puntuación. Este es un valor predeterminado.

Puntuación (Continuidad corregida)
Muestra la estadística de prueba Z de puntuación con continuidad corregida.

Wald
Muestra la estadístico de prueba Z de Wald.

Wald (Continuidad corregida)
Muestra la estadístico de prueba Z de Wald con continuidad corregida.

Valor de prueba
Especifica un valor de prueba entre 0 y 1. El valor predeterminado es 0.5.

Obtención de pruebas de proporciones de una muestra

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de una muestra...**
2. En el diálogo Proporciones de una muestra, pulse **Pruebas**.
3. Seleccione una o más de las pruebas disponibles.

Proporciones de una muestra: Valores perdidos

El diálogo Valores perdidos proporciona opciones para tratar los valores perdidos.

Falta el ámbito de datos

Excluir casos según análisis
Indica la inclusión de todos los casos con datos suficientes en las variables utilizadas en cada análisis concreto. Ésta es el ajuste predeterminado.

Excluir casos según lista
Indica la inclusión de todos los casos con datos suficientes en todas las variables utilizadas en todos los análisis.

Valores perdidos de usuario
Excluir trata los valores perdidos de usuario como ausentes. **Incluir** no tiene en cuenta las designaciones de valor perdido y trata los valores perdidos de usuario como válidos.

Definición de valores para los valores perdidos para proporciones de una muestra

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de una muestra...**

2. En el diálogo Proporciones de una muestra, pulse **Valores perdidos**.
3. Seleccione los valores deseados para los valores perdidos.

Proporciones de muestras emparejadas

El procedimiento Proporciones de muestras emparejadas proporciona pruebas e intervalos de confianza para la diferencia entre dos proporciones binomiales relacionadas o emparejadas. Se presupone que los datos proceden de una única muestra aleatoria, y cada prueba de hipótesis o intervalo de confianza es una prueba independiente o un intervalo individual. La salida incluye las proporciones observadas, estimaciones de las diferencias en las proporciones de población, errores estándar asintóticos de las diferencias de población bajo hipótesis nulas y alternativas, estadísticos de prueba especificados con probabilidades bilaterales, e intervalos de confianza especificados para las diferencias en las proporciones.

Ejemplo

Estadísticos
Agresti-Min, Bonett-Price, Newcombe, Wald, Wald (continuidad corregida), Binomial exacto, Binomial ajustado mid-p, McNemar, McNemar (continuidad corregida).

Consideraciones sobre los datos

Datos

- Se requiere una lista de variables que contenga al menos dos variables.
- Si se especifica una única lista de variables, cada miembro de la lista está emparejado con todos los demás miembros de la lista.

Supuestos

- Si se separan dos listas de variables con **WITH** sin la palabra clave (**PAIRED**), cada miembro de la primera lista se empareja con cada uno de los miembros de la segunda lista.
- Si dos listas de variables están separadas por **WITH** y la segunda lista es seguida por (**PAIRED**), se emparejan los miembros de las dos listas en orden: el primer miembro de la primera lista se empareja con el primer miembro de la segunda lista, los segundos miembros de cada lista se emparejan, etc. Se ignoran las variables no coincidentes y se emite un mensaje de aviso.

Obtención de pruebas de proporciones de muestras emparejadas

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de muestras emparejadas...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables de contraste cuantitativas.
3. Si lo desea, puede:
   - Seleccione valores de criterios de éxito en la sección **Definir éxito**:

     Último valor
     Se utiliza el valor más alto o último entre los valores distintos ordenados de los datos. Esto se aplica a variables numéricas o de cadena. Ésta es el ajuste predeterminado.

     Primer valor
     Se utiliza el valor más bajo o primero entre los valores distintos ordenados de los datos. Esto se aplica a variables numéricas o de cadena.

     Valores
     Uno o más valores específicos entre paréntesis. Si hay varios valores, se deben separar por espacios. Esto se aplica a variables numéricas o de cadena. Los valores de variable de cadena se deben delimitar entre comillas simples.

Punto medio
Los valores que están en la mitad o por encima del rango de valores observados de los datos. Esto se aplica únicamente a datos numéricos.

Punto de corte
Valores que están en un valor especificado o por encima de él. Esto se aplica únicamente a datos numéricos.

- Haga clic en **Intervalos de confianza...** para especificar qué tipos de intervalos de confianza se muestran, o para suprimir todos los intervalos de confianza.
- Haga clic en **Pruebas...** para especificar qué tipos de estadísticas de pruebas se muestran, o para suprimir todas las pruebas.
- Haga clic en **Valores perdidos...** para controlar el tratamiento de los datos que faltan.
- Haga clic en **Bootstrap...** para derivar estimaciones robustas de errores estándar e intervalos de confianza para estimaciones como la media, la mediana, la proporción, la odds ratio, el coeficiente de correlación o el coeficiente de regresión.

4. Pulse **Aceptar**.

Proporciones de muestras emparejadas: Intervalos de confianza

El diálogo Intervalos de confianza proporciona opciones para especificar el nivel de cobertura y para seleccionar qué tipos de intervalos de confianza se muestran.

Nivel de cobertura
Especifica el porcentaje del intervalo de confianza. Se debe especificar un valor numérico en el rango (0,100). 95 es el valor predeterminado.

Tipos de intervalos
Proporciona opciones para especificar qué tipos de intervalos de confianza se muestran. Las opciones disponibles son:

- Agresti-Min
- Bonett-Price
- Newcombe
- Wald
- Wald (Continuidad corregida)

Especificación de intervalos de confianza para proporciones de muestras emparejadas

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de muestras emparejadas...**

2. Pulse **Intervalos de confianza** para especificar qué tipos de intervalos de confianza se muestran, o para suprimir todos los intervalos de confianza.

Proporciones de muestras emparejadas: Pruebas

El diálogo Pruebas proporciona opciones para especificar qué tipos de estadísticos de prueba se muestran.

Todos
Se muestran todas las estadísticos de prueba en los resultados.

Ninguna
No se muestran estadísticos de prueba en los resultados.

Binomial exacto
Muestra probabilidades binomiales exactas.

Binomial ajustado mid-p
Muestra probabilidades binomiales ajustadas mid-p. Este es un valor predeterminado.

McNemar
Muestra la estadístico de prueba Z de McNemar. Este es un valor predeterminado.

McNemar (Continuidad corregida)
Muestra la estadístico de prueba Z de McNemar con continuidad corregida.

Wald
Muestra la estadístico de prueba Z de Wald.

Wald (Continuidad corregida)
Muestra la estadístico de prueba Z de Wald con continuidad corregida.

Obtención de pruebas de proporciones de muestras emparejadas

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de muestras emparejadas...**
2. En el diálogo Proporciones de muestras emparejadas, pulse **Pruebas**.
3. Seleccione una o más de las pruebas disponibles.

Proporciones de muestras emparejadas: Valores perdidos

El diálogo Valores perdidos proporciona opciones para tratar los valores perdidos.

Falta el ámbito de datos

Excluir casos según análisis
Indica la inclusión de todos los casos con datos suficientes en las variables utilizadas en cada análisis concreto. Ésta es el ajuste predeterminado.

Excluir casos según lista
Indica la inclusión de todos los casos con datos suficientes en todas las variables utilizadas en todos los análisis.

Valores perdidos de usuario
Excluir trata los valores perdidos de usuario como ausentes. **Incluir** no tiene en cuenta las designaciones de valor perdido y trata los valores perdidos de usuario como válidos.

Definición de valores para los valores perdidos para proporciones de muestras emparejadas

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de muestras emparejadas...**
2. En el diálogo Proporciones de muestras emparejadas, pulse **Valores perdidos**.
3. Seleccione los valores deseados para los valores perdidos.

Proporciones de muestras independientes

El procedimiento Proporciones de muestras independientes proporciona pruebas e intervalos de confianza para la diferencia entre dos proporciones binomiales independientes. Se presupone que los datos proceden de una única muestra aleatoria, y cada prueba de hipótesis o intervalo de confianza es una prueba independiente o un intervalo individual. La salida incluye las proporciones observadas, estimaciones de las diferencias en las proporciones de población, errores estándar asintóticos de las diferencias de población bajo hipótesis nulas y alternativas, estadísticos de prueba especificados con probabilidades bilaterales, e intervalos de confianza especificados para las diferencias en las proporciones.

Ejemplo

Estadísticos

Agresti-Min, Bonett-Price, Newcombe, Wald, Wald (continuidad corregida), Binomial exacto, Binomial ajustado mid-p, McNemar, McNemar (continuidad corregida).

Consideraciones sobre los datos

Datos

- Se necesita al menos una variable dependiente y una variable única para identificar los dos grupos que se van a comparar.
- La variable de agrupación puede ser numérica o de cadena.

Supuestos

Obtención de pruebas de proporciones de muestras independientes

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de muestras independientes...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables de contraste cuantitativas.
3. Seleccione una **Variable de agrupación** individual que identifique los dos grupos que se van a comparar.
4. Si lo desea, especifique valores para la **Variable de agrupación** seleccionada.
   - Cuando se selecciona **Valores**, puede especificar dos valores numéricos o de cadena entre paréntesis como los valores que se van a comparar. Los valores de cadena deben indicarse entre comillas simples. Los casos que tengan otros valores no se tienen en cuenta.
   - **Punto medio** solo se aplica a variables numéricas. Los casos que estén en el punto medio de la distribución de la variable de agrupación o por encima de él se asignarán al segundo grupo, mientras que los casos que estén por debajo del punto medio se asignarán al primer grupo.
   - **Punto de corte** solo se aplica a las variables numéricas y permite la especificación con paréntesis de un único valor numérico. Los casos que estén en el punto de corte de la variable de agrupación o por encima de él se asignarán al segundo grupo, mientras que los casos que estén por debajo del punto de corte se asignarán al primer grupo.
5. Si lo desea, puede:
   - Seleccione valores de criterios de éxito en la sección **Definir éxito**:

     Último valor
     Se utiliza el valor más alto o último entre los valores distintos ordenados de los datos. Esto se aplica a variables numéricas o de cadena. Ésta es el ajuste predeterminado.

     Primer valor
     Se utiliza el valor más bajo o primero entre los valores distintos ordenados de los datos. Esto se aplica a variables numéricas o de cadena.

     Valores
     Uno o más valores específicos entre paréntesis. Si hay varios valores, se deben separar por espacios. Esto se aplica a variables numéricas o de cadena. Los valores de variable de cadena se deben delimitar entre comillas simples.

     Punto medio
     Los valores que están en la mitad o por encima del rango de valores observados de los datos. Esto se aplica únicamente a datos numéricos.

     Punto de corte
     Valores que están en un valor especificado o por encima de él. Esto se aplica únicamente a datos numéricos.

- Haga clic en **Intervalos de confianza...** para especificar qué tipos de intervalos de confianza se muestran, o para suprimir todos los intervalos de confianza.
- Haga clic en **Pruebas...** para especificar qué tipos de estadísticas de pruebas se muestran, o para suprimir todas las pruebas.
- Haga clic en **Valores perdidos...** para controlar el tratamiento de los datos que faltan.
- Haga clic en **Bootstrap...** para derivar estimaciones robustas de errores estándar e intervalos de confianza para estimaciones como la media, la mediana, la proporción, la odds ratio, el coeficiente de correlación o el coeficiente de regresión.

6. Pulse **Aceptar**.

Proporciones de muestras independientes: Intervalos de confianza

El diálogo Intervalos de confianza proporciona opciones para especificar el nivel de cobertura y para seleccionar qué tipos de intervalos de confianza se muestran.

Nivel de cobertura
Especifica el porcentaje del intervalo de confianza. Se debe especificar un valor numérico en el rango (0,100). 95 es el valor predeterminado.

Tipos de intervalos
Proporciona opciones para especificar qué tipos de intervalos de confianza se muestran. Las opciones disponibles son:

- Agresti-Caffo
- Brown-Li-Jeffreys
- Hauck-Anderson
- Newcombe
- Newcombe (Continuidad corregida)
- Wald
- Wald (Continuidad corregida)

Especificación de intervalos de confianza para proporciones de muestras independientes

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de muestras independientes...**

2. Pulse **Intervalos de confianza** para especificar qué tipos de intervalos de confianza se muestran, o para suprimir todos los intervalos de confianza.

Proporciones de muestras independientes: Pruebas

El diálogo Pruebas proporciona opciones para especificar qué tipos de estadísticos de prueba se muestran.

Todos
Se muestran todas las estadísticos de prueba en los resultados.

Ninguna
No se muestran estadísticos de prueba en los resultados.

Hauck-Anderson
Muestra la estadística de prueba Z de Hauck-Anderson.

Wald
Muestra la estadístico de prueba Z de Wald.

Wald (Continuidad corregida)
Muestra la estadístico de prueba Z de Wald con continuidad corregida.

Wald H0
Muestra el estadístico de prueba Z de Wald utilizando estimaciones de varianza por debajo de H_0.

Wald H0 (Continuidad corregida)
Muestra el estadístico de prueba Z de Wald con continuidad corregida utilizando estimaciones de varianza por debajo de H_0.

Obtención de pruebas de proporciones de muestras independientes

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de muestras independientes...**
2. En el diálogo Proporciones de muestras independientes, pulse **Pruebas**.
3. Seleccione una o más de las pruebas disponibles.

Proporciones de muestras independientes: Valores perdidos

El diálogo Valores perdidos proporciona opciones para tratar los valores perdidos.

Falta el ámbito de datos

Excluir casos según análisis
Indica la inclusión de todos los casos con datos suficientes en las variables utilizadas en cada análisis concreto. Ésta es el ajuste predeterminado.

Excluir casos según lista
Indica la inclusión de todos los casos con datos suficientes en todas las variables utilizadas en todos los análisis.

Valores perdidos de usuario
Excluir trata los valores perdidos de usuario como ausentes. **Incluir** no tiene en cuenta las designaciones de valor perdido y trata los valores perdidos de usuario como válidos.

Definición de valores para los valores perdidos para proporciones de muestras independientes

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Proporciones de muestras independientes...**
2. En el diálogo Proporciones de muestras independientes, pulse **Valores perdidos**.
3. Seleccione los valores deseados para los valores perdidos.

Pruebas T

Pruebas T

Hay tres tipos de pruebas *t*:

Prueba T de muestras independientes (prueba T para dos muestras). Compara las medias de una variable para dos grupos de casos. Se ofrecen estadísticos descriptivos para cada grupo y la prueba de Levene sobre la igualdad de las varianzas, así como valores *t* de igualdad de varianzas y varianzas desiguales y un intervalo de confianza al 95% para la diferencia entre las medias.

Prueba T de muestras emparejadas (prueba T dependiente). Compara las medias de dos variables en un solo grupo. Esta prueba también se utiliza para pares coincidentes o diseños de estudio de control de casos. El resultado incluye estadísticos descriptivos de las variables que se van a contrastar, la correlación entre ellas, estadísticos descriptivos de las diferencias emparejadas, la prueba *t* y un intervalo de confianza al 95%.

Prueba t para una muestra. Compara la media de una variable con un valor conocido o hipotetizado. Se muestran estadísticos descriptivos para las variables de contraste junto con la prueba *t*. Un intervalo

de confianza del 95% para la diferencia entre la media de la variable de prueba y el valor de prueba hipotetizado forma parte del resultado predeterminado.

Prueba T para muestras independientes

El procedimiento de prueba T de muestras independientes compara medias para dos grupos de casos y automatiza el cálculo del tamaño de efecto de prueba *t*. Lo ideal para esta prueba es asignar aleatoriamente a los sujetos a dos grupos, de modo que cualquier diferencia en la respuesta se deba al tratamiento (o a la falta de tratamiento) y no a otros factores. Este caso no ocurre si se comparan los ingresos medios para hombres y mujeres. El sexo de una persona no se asigna aleatoriamente. En estas situaciones, debe asegurarse de que las diferencias en otros factores no enmascaren o resalten una diferencia significativa entre las medias. Las diferencias de ingresos medios pueden estar sometidas a la influencia de factores como los estudios (y no solamente el sexo).

Ejemplo

Se asigna aleatoriamente un grupo de pacientes con hipertensión arterial a un grupo con placebo y otro con tratamiento. Los sujetos con placebo reciben una pastilla inactiva y los sujetos con tratamiento reciben un nuevo medicamento del cual se espera que reduzca la tensión arterial. Después de tratar a los sujetos durante dos meses, se utiliza la prueba *t* para dos muestras para comparar la tensión arterial media del grupo con placebo y del grupo con tratamiento. Cada paciente se mide una sola vez y pertenece a un solo grupo.

Estadísticos

Para cada variable: tamaño de la muestra, media, desviación estándar, error estándar de la media y la estimación del tamaño de efecto para la prueba *t*. Para la diferencia entre las medias: media, error estándar e intervalo de confianza (puede especificar el nivel de confianza). Pruebas: prueba de Levene sobre la igualdad de varianzas y pruebas *t* de varianzas combinadas y separadas sobre la igualdad de las medias.

Consideraciones sobre los datos

Datos

Los valores de la variable cuantitativa de interés se hallan en una única columna del archivo de datos. El procedimiento utiliza una variable de agrupación con dos valores para separar los casos en dos grupos. La variable de agrupación puede ser numérica (valores como 1 y 2, o 6,25 y 12,5) o de cadena corta (como *sí* y *no*). También puede usar una variable cuantitativa, como la *edad*, para dividir los casos en dos grupos especificando un punto de corte (el punto de corte 21 divide la *edad* en un grupo de menos de 21 años y otro de más de 21).

Supuestos

Para la prueba *t* de igualdad de varianzas, las observaciones deben ser muestras aleatorias independientes de distribuciones normales con la misma varianza de población. Para la prueba *t* de varianzas desiguales, las observaciones deben ser muestras aleatorias independientes de distribuciones normales. La prueba *t* para dos muestras es bastante robusta a las desviaciones de la normalidad. Al contrastar las distribuciones gráficamente, compruebe que son simétricas y que no contienen valores atípicos.

Obtención de una prueba T de muestras independientes

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Prueba T de muestras independientes...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables de contraste cuantitativas. Se calcula una prueba *t* para cada variable.
3. Seleccione una sola variable de agrupación y pulse en **Definir grupos** para especificar dos códigos para los grupos que desee comparar.
4. Los siguientes ajustes son opcionales.

- Seleccione **Estimar tamaños de efecto** para controlar la estimación del tamaño de efecto de prueba *t*.
- Seleccione **Prueba de homogeneidad de la varianza** para calcular el estadístico de Levene para comprobar la igualdad de las varianzas de los grupos. Puede obtener los resultados en una tabla aparte.
- Pulse **Opciones** para controlar el tratamiento de los datos perdidos y el nivel del intervalo de confianza.
- Pulse **Programa de arranque** para derivar estimaciones robustas de errores estándar e intervalos de confianza para estimaciones como, por ejemplo, la media, mediana, proporción, razón de probabilidad, coeficiente de correlación o coeficiente de regresión.

Definición de grupos de prueba T de muestras independientes

Para las variables de agrupación numéricas, defina los dos grupos para la prueba *t* especificando dos valores o un punto de corte:

- **Usar valores especificados.** Especifique un valor para el grupo 1 y otro valor para el grupo 2. Los casos con cualquier otro valor se excluyen del análisis. Los números no tienen que ser enteros (por ejemplo, 6,25 y 12,5 son válidos).
- **Punto de corte.** Escriba un número que divida los valores de la variable de agrupación en dos conjuntos. Todos los casos con valores menores que el punto de corte forman un grupo y los casos con valores mayores o iguales que el punto de corte forman el otro grupo.

Para las variables de agrupación de series, especifique una serie para el grupo 1 y otro valor para el grupo 2, como por ejemplo *Sí* y *no*. Los casos con otras series se excluyen del análisis.

Prueba T para muestras independientes: Opciones

Intervalo de confianza. De forma predeterminada, se muestra un intervalo de confianza al 95% para la diferencia entre las medias. Introduzca un valor entre 1 y 99 para solicitar otro nivel de confianza.

Valores perdidos. Si ha probado varias variables y se han perdido los datos de una o más de ellas, puede indicar al procedimiento qué casos desea incluir (o excluir).

- **Excluir casos según análisis.** Cada prueba *t* utiliza todos los casos que tienen datos válidos para las variables contrastadas. Los tamaños muestrales pueden variar de una prueba a otra.
- **Excluir casos según lista.** Cada prueba *t* utiliza sólo aquellos casos que contienen datos válidos para todas las variables utilizadas en las pruebas *t* solicitadas. El tamaño de la muestra es constante en todas las pruebas.

Prueba T de muestras emparejadas

El procedimiento Prueba T de muestras emparejadas compara las medias de dos variables de un solo grupo. El procedimiento calcula las diferencias entre los valores de las dos variables de cada caso y contrasta si la media difiere de 0. El procedimiento también automatiza el cálculo de tamaño de efecto de prueba de *t*.

Ejemplo

En un estudio sobre la hipertensión sanguínea, se toma la tensión a todos los pacientes al comienzo del estudio, se les aplica un tratamiento y se les toma la tensión otra vez. De esta manera, a cada sujeto le corresponden dos medidas, normalmente denominadas medidas *pre* y *post*. Un diseño alternativo para el que se utiliza esta prueba consiste en un estudio de pares relacionados o un estudio de control de casos en el que cada registro en el archivo de datos contiene la respuesta del paciente y de su sujeto de control correspondiente. En un estudio sobre la tensión sanguínea, pueden emparejarse pacientes y controles por edad (un paciente de 75 años con un miembro del grupo de control de 75 años).

Estadísticos

Para cada variable: media, tamaño de la muestra, desviación estándar y error estándar de la media. Para cada par de variables: correlación, diferencia promedio entre las medias, la prueba *t*, el intervalo

de confianza para la diferencia entre las medias (puede especificarse el nivel de confianza) y la estimación del tamaño de efecto de la prueba *t*. Desviación estándar y error estándar de la diferencia entre las medias.

Consideraciones sobre los datos

Datos

Especifique dos variables cuantitativas (nivel de medición de intervalo o de razón) para cada prueba de pares. En un estudio de pares relacionados o de control de casos, la respuesta de cada sujeto de la prueba y su sujeto de control correspondiente deberán hallarse en el mismo caso en el archivo de datos.

Supuestos

Las observaciones de cada par deben hacerse en las mismas condiciones. Las diferencias entre las medias deben estar normalmente distribuidas. Las varianzas de cada variable pueden ser iguales o desiguales.

Obtener una prueba T de muestras emparejadas

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Prueba T de muestras emparejadas...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione uno o más pares de variables.
3. Opcionalmente, puede cambiar/seleccionar una opción **Estimar tamaños de efecto**. Los valores controlan cómo se calcula el estandarizador en la estimación de la corrección *d* de Cohen y la corrección de Hedges para cada par de variables.

   Desviación estándar de la diferencia

   El denominador utilizado en la estimación del tamaño de efecto. La *d* de Cohen utiliza la desviación estándar de la diferencia de medias. La corrección de Hedges utiliza la desviación estándar de la muestra de la diferencia media ajustada por un factor de corrección.

   Desviación estándar corregida de la diferencia

   El denominador utilizado en la estimación del tamaño de efecto. La *d* de Cohen utiliza la desviación estándar de la muestra de la diferencia media ajustada por la correlación entre las medidas ajustada por la correlación entre las medidas. La corrección de Hedges utiliza la desviación estándar de la muestra de la diferencia media ajustada por la correlación entre las medidas, más un factor de corrección.

   Promedio de varianzas

   El denominador utilizado en la estimación del tamaño de efecto. La *d* de Cohen utiliza la raíz cuadrada de la varianza promedio de las medidas. La corrección de Hedges utiliza la raíz cuadrada de la varianza promedio de las medidas, más un factor de corrección.

4. Si lo desea, puede:

   - Seleccione **Estimar tamaños de efecto** para controlar la estimación del tamaño de efecto de prueba *t*. Cuando se selecciona el valor, puede controlar cómo se calcula el estandarizador en la estimación de la corrección *d* de Cohen y la corrección de Hedges para cada par de variables.
   - Pulse **Opciones** para controlar el tratamiento de los datos perdidos y el nivel del intervalo de confianza.
   - Pulse **Programa de arranque** para derivar estimaciones robustas de errores estándar e intervalos de confianza para estimaciones como, por ejemplo, la media, mediana, proporción, razón de probabilidad, coeficiente de correlación o coeficiente de regresión.

Prueba T para muestras relacionadas: Opciones

Intervalo de confianza. De forma predeterminada, se muestra un intervalo de confianza al 95% para la diferencia entre las medias. Introduzca un valor entre 1 y 99 para solicitar otro nivel de confianza.

Valores perdidos. Si ha probado varias variables y se han perdido los datos de una o más de ellas, puede indicar al procedimiento qué casos desea incluir (o excluir):

- **Excluir casos según análisis.** Cada prueba *t* utilizará todos los casos que contienen datos válidos para la pareja de variables contrastadas. Los tamaños muestrales pueden variar de una prueba a otra.
- **Excluir casos según lista.** Cada prueba *t* utilizará únicamente los casos que contengan datos válidos para todas las parejas de variables contrastadas. El tamaño de la muestra es constante en todas las pruebas.

Características adicionales del comando T TEST

La sintaxis de comandos también le permite:

- Producir pruebas t tanto de una sola muestra como de muestras independientes ejecutando un solo comando.
- Contrastar una variable con todas las variables de una lista, en una prueba relacionada (mediante el subcomando PAIRS).
- Controlar la estimación del tamaño del efecto de prueba *t* (con el subcomando ES).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Prueba T para una muestra

El procedimiento Prueba T de una muestra prueba si la media de una única variable difiere de una constante especificada y automatiza el cálculo del tamaño de efecto de prueba *t*.

Ejemplos

Un investigador podría querer probar si la puntuación promedio de IQ para un grupo de estudiantes difiere de 100. O un fabricante de cereales puede tomar una muestra de cajas de la línea de producción y comprobar si el peso medio de las muestras difiere de 1,3 libras en el nivel de confianza del 95%.

Estadísticos

Para cada variable de prueba: media, desviación estándar, error estándar de la media, y la estimación del tamaño de efecto para la prueba *t*. La diferencia promedio entre cada valor de los datos y el valor de contraste hipotetizado, una prueba *t* que contrasta que esta diferencia es 0 y un intervalo de confianza para la diferencia promedio (para el que puede especificarse el nivel de confianza).

Consideraciones sobre los datos

Datos

Para contrastar los valores de una variable cuantitativa con un valor de contraste hipotetizado, elija una variable cuantitativa e introduzca un valor de contraste hipotetizado.

Supuestos

Esta prueba asume que los datos están normalmente distribuidos; sin embargo, esta prueba es bastante robusto frente a las desviaciones de la normalidad.

Obtención de una prueba T para una muestra

1. En los menús seleccione:

   Analizar > **Comparar medias** > **Prueba T de una muestra...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables para contrastarlas con el mismo valor hipotetizado.

3. Introduzca un valor de contraste numérico para compararlo con cada media muestral.
4. Si lo desea, puede:
   - Seleccione **Estimar tamaños de efecto** para controlar la estimación del tamaño de efecto de prueba *t*.
   - Pulse **Opciones** para controlar el tratamiento de los datos perdidos y el nivel del intervalo de confianza.

Prueba T para una muestra: Opciones

Intervalo de confianza. De forma predeterminada, se muestra un intervalo de confianza al 95% para la diferencia entre la media y el valor de contraste hipotetizado. Introduzca un valor entre 1 y 99 para solicitar otro nivel de confianza.

Valores perdidos. Si ha probado varias variables y se han perdido los datos de una o más de ellas, puede indicar al procedimiento qué casos desea incluir (o excluir).

- **Excluir casos según análisis.** Cada prueba *t* utiliza todos los casos que tienen datos válidos para la variable contrastada. Los tamaños muestrales pueden variar de una prueba a otra.
- **Excluir casos según lista.** Cada prueba *t* utiliza sólo aquellos casos que contienen datos válidos para todas las variables utilizadas en las pruebas *t* solicitadas. El tamaño de la muestra es constante en todas las pruebas.

Características adicionales del comando T TEST

La sintaxis de comandos también le permite:

- Producir pruebas t tanto de una sola muestra como de muestras independientes ejecutando un solo comando.
- Contrastar una variable con todas las variables de una lista, en una prueba relacionada (mediante el subcomando PAIRS).
- Controlar la estimación del tamaño del efecto de prueba *t* (con el subcomando ES).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Características adicionales del comando T TEST

La sintaxis de comandos también le permite:

- Producir pruebas t tanto de una sola muestra como de muestras independientes ejecutando un solo comando.
- Contrastar una variable con todas las variables de una lista, en una prueba relacionada (mediante el subcomando PAIRS).
- Controlar la estimación del tamaño del efecto de prueba *t* (con el subcomando ES).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

ANOVA de un factor

El procedimiento ANOVA de un factor genera un análisis de varianza de un factor para una variable dependiente cuantitativa respecto a una única variable de factor (la variable independiente) y estima el tamaño de efecto en ANOVA de un factor. El análisis de varianza se utiliza para contrastar la hipótesis de que varias medias son iguales. Esta técnica es una extensión de la prueba *t* para dos muestras.

Además de determinar que existen diferencias entre las medias, es posible que desee saber qué medias difieren. Existen dos tipos de contrastes para comparar medias: a priori y post hoc. Los contrastes a priori se plantean *antes* de ejecutar el experimento y las pruebas post hoc se realizan *después* de haber llevado a cabo el experimento. También puede contrastar las tendencias existentes a través de las categorías.

Ejemplo

Las rosquillas absorben diferentes cantidades de grasa cuando se fríen. Se plantea un experimento utilizando tres tipos de grasas: aceite de cacahuete, aceite de maíz y manteca de cerdo. El aceite de cacahuete y el aceite de maíz son grasas no saturadas y la manteca es una grasa saturada. Además de determinar si la cantidad de grasa absorbida depende del tipo de grasa utilizada, también se podría preparar un contraste a priori para determinar si la cantidad de absorción de la grasa difiere para las grasas saturadas y las no saturadas.

Estadísticos

Para cada grupo: número de casos, media, desviación estándar, error estándar de la media, mínimo, máximo, intervalo de confianza al 95% para la media y la estimación del tamaño de efecto en ANOVA de un factor. Pruebas de Levene sobre la homogeneidad de varianzas, tabla de análisis de varianza y contrastes robustos de igualdad de medias para cada variable dependiente, contrastes a priori especificados por el usuario y las pruebas de rango y de comparaciones múltiples post hoc: Bonferroni, Sidak, diferencia honestamente significativa de Tukey, GT2 de Hochberg, Gabriel, Dunnett, prueba *F* de Ryan-Einot-Gabriel-Welsch, (R-E-G-W *F*), prueba de rango de Ryan-Einot-Gabriel-Welsch (R-E-G-W *Q*), T2 de Tamhane, T3 de Dunnett, Games-Howell, *C*, de Dunnett, prueba de rango múltiple de Duncan, Student-Newman-Keuls (S-N-K), *b* de Tukey, Waller-Duncan, Scheffé y diferencia menos significativa.

Consideraciones sobre los datos

Datos

La variable dependiente debe ser cuantitativa (nivel de medición de intervalo).

Supuestos

Cada grupo es una muestra aleatoria independiente procedente de una población normal. El análisis de varianza es robusto a las desviaciones de la normalidad, aunque los datos deberán ser simétricos. Los grupos deben proceder de poblaciones con varianzas iguales. Para contrastar este supuesto, utilice la prueba de Levene de homogeneidad de varianzas.

Cómo obtener un análisis de varianza de un factor

1. En los menús seleccione:

   Analizar > **Comparar medias** > **ANOVA unidireccional...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables dependientes.
3. Seleccione una sola variable de factor independiente.

Si lo desea, puede:

- Seleccione **Estimar tamaño de efecto para pruebas generales** para controlar el cálculo del tamaño de efecto para la prueba global. Cuando se selecciona, la tabla “Tamaños de efecto ANOVA” se visualiza en la salida.
- Pulse **Contrastes** para dividir las sumas entre grupos de cuadrados en componentes de tendencia o especificar contrastes a priori.
- Pulse **Post Hoc** para utilizar las pruebas de rango post hoc y comparaciones múltiples por parejas para determinar qué medidas difieren.
- Pulse **Opciones** para controlar el tratamiento de los datos perdidos y el nivel del intervalo de confianza.
- Pulse **Programa de arranque** para derivar estimaciones robustas de errores estándar e intervalos de confianza para estimaciones como, por ejemplo, la media, mediana, proporción, razón de probabilidad, coeficiente de correlación o coeficiente de regresión.

ANOVA de un factor: Contrastes

Puede dividir las sumas de cuadrados inter-grupos en componentes de tendencia o especificar contrastes a priori.

Polinómica
Divide las sumas de cuadrados inter-grupos en componentes de tendencia. Puede contrastar la existencia de tendencia en la variable dependiente a través de los niveles ordenados de la variable de factor. Por ejemplo, podría contrastar si existe una tendencia lineal (creciente o decreciente) en el salario, a través de los niveles ordenados de la titulación mayor obtenida.

- **Orden.** Se puede elegir un orden polinómico 1º, 2º, 3º, 4º o 5º.

Coeficientes
Contrastes a priori especificados por el usuario que serán contrastados mediante el estadístico *t*. Introduzca un coeficiente para cada grupo (categoría) de la variable factor y pulse **Añadir** después de cada entrada. Cada nuevo valor se añade al final de la lista de coeficientes. Para especificar conjuntos de contrastes adicionales, pulse **Siguiente**. Utilice **Siguiente** y **Anterior** para desplazarse por los conjuntos de contrastes.

Estimar tamaño de efecto para contrastes
Controla el cálculo del tamaño de efecto de la prueba general. Cuando este parámetro está habilitado, al menos una de las opciones siguientes debe estar seleccionada para poder calcular los tamaños de efecto. Este valor se habilita cuando se especifica al menos un contraste y da como resultado una tabla de tamaños de efecto ANOVA en los resultados.

Utilizar desviación estándar agrupada para todos los grupos como estandarizador
Utiliza la desviación estándar agrupada para todos los grupos como estandarizador al estimar el tamaño de efecto. Este es el valor predeterminado y está disponible cuando **Tamaño de efecto de estimación para contrastes** está seleccionado.

Utilizar desviación estándar agrupada para aquellos grupos implicados en el contraste como estandarizador
Utiliza la desviación estándar agrupada para los grupos implicados en el contraste como estandarizador. Esta parámetro está disponible cuando **Tamaño de efecto de estimación para contrastes** está seleccionado.

El orden de los coeficientes es importante porque se corresponde con el orden ascendente de los valores de las categorías de la variable de factor. El primer coeficiente en la lista se corresponde con el menor de los valores de grupo en la variable de factor y el último coeficiente se corresponde con el valor más alto. Por ejemplo, si existen seis categorías en la variable factor, los coeficientes -1, 0, 0, 0, 0,5 y 0,5 contrastan el primer grupo con los grupos quinto y sexto. Para la mayor parte de las aplicaciones, los coeficientes deben sumar 0. Los conjuntos que no suman 0 también se pueden utilizar, pero se visualiza un mensaje de aviso.

ANOVA de un factor: Contrastes post hoc

Una vez que se ha determinado que existen diferencias entre las medias, las pruebas de rango post hoc y las comparaciones múltiples por parejas permiten determinar qué medias difieren. Las pruebas de rango identifican subconjuntos homogéneos de medias que no se diferencian entre sí. Las comparaciones múltiples por parejas contrastan la diferencia entre cada pareja de medias y generan una matriz donde los asteriscos indican las medias de grupo significativamente diferentes a un nivel alfa de 0,05.

Asumiendo varianzas iguales

La prueba de la diferencia honestamente significativa de Tukey, la GT2 de Hochberg, la prueba de Gabriel y la prueba de Scheffé son pruebas de comparaciones múltiples y pruebas de rango. Otras pruebas de rango disponibles son la *b* de Tukey, S-N-K (Student-Newman-Keuls), Duncan, R-E-G-W *F* (prueba *F* de Ryan-Einot-Gabriel-Welsch), R-E-G-W *Q* (prueba de rango de Ryan-Einot-Gabriel-Welsch) y Waller-Duncan. Las pruebas de comparaciones múltiples disponibles son Bonferroni, Diferencia honestamente significativa de Tukey, Sidak, Gabriel, Hochberg, Dunnett, Scheffé y DMS (diferencia menos significativa).

- *LSD*. Utiliza pruebas t para realizar todas las comparaciones por parejas entre medias de grupo. La tasa de error no se corrige para realizar múltiples comparaciones.
- *Bonferroni*. Utiliza las pruebas *t* para realizar comparaciones por parejas entre las medias de los grupos, pero controla la tasa de error global estableciendo que la tasa de error de cada prueba sea igual a la tasa de error por experimento dividida entre el número total de contrastes. Así, se corrige el nivel de significación observado por el hecho de que se están realizando múltiples comparaciones.
- *Sidak*. Prueba de comparación múltiple por parejas basada en un estadístico t. La prueba de Sidak corrige el nivel de significación para las comparaciones múltiples y da lugar a límites más estrechos que los de Bonferroni.
- *Scheffe*. Realiza comparaciones por parejas conjuntas simultáneas para todas las combinaciones posibles por parejas de medias. Utiliza la distribución muestral F. Puede utilizarse para examinar todas las combinaciones lineales de grupos de medias posibles, no sólo las comparaciones por parejas.
- *R-E-G-W F*. Procedimiento múltiple por pasos (por tamaño de las distancias) de Ryan-Einot-Gabriel-Welsch que se basa en una prueba F.
- *R-E-G-W Q*. Procedimiento múltiple por pasos (por tamaño de las distancias) de Ryan-Einot-Gabriel-Welsch que se basa en el rango estudentizado.
- *S-N-K*. Realiza todas las comparaciones por parejas entre las medias utilizando la distribución de rango estudentizado. Con tamaños de muestras iguales, también compara pares de medias dentro de subconjuntos homogéneos utilizando un procedimiento por pasos Las medias se ordenan de mayor a menor y se comparan primero las diferencias más extremas.
- *Tukey*. Utiliza el estadístico del rango estudentizado para realizar todas las comparaciones por parejas entre grupos. Establece la tasa de error por experimento como la tasa de error para el conjunto de todas las comparaciones por parejas.
- *Tukey-b*. Utiliza la distribución de rango estudentizado para realizar comparaciones por parejas entre grupos. El valor crítico es el promedio de los valores correspondientes a la diferencia honestamente significativa de Tukey y al método de Student-Newman-Keuls.
- *Duncan*. Realiza comparaciones por parejas utilizando un orden por pasos de comparaciones idéntico al orden utilizado por la prueba de Student-Newman-Keuls, pero establece un nivel de protección para la tasa de error para la colección de pruebas, en lugar de una tasa de error para pruebas individuales. Utiliza el estadístico del rango estudentizado.
- *GT2 de Hochberg.* Prueba de rango y comparación múltiple que utiliza el módulo máximo estudentizado. Es similar a la prueba de la diferencia honestamente significativa de Tukey.
- *Gabriel*. Prueba de comparación por parejas que ha utilizado el módulo máximo estudentizado y que generalmente es más potente que el GT2 de Hochberg cuando los tamaños de celda son desiguales. La prueba de Gabriel se puede convertir en liberal cuando los tamaños de las casillas varían mucho.
- *Waller-Duncan*. Prueba de comparaciones múltiples basada en un estadístico t. Utiliza la aproximación Bayesiana.
- *Dunnett*. Prueba *t* de comparaciones múltiples por parejas que compara un conjunto de tratamientos respecto a una única medida de control.La última categoría es la categoría de control predeterminada. Si lo desea, puede seleccionar la primera categoría. Pruebas **bilaterales** que indican que la media en cualquier nivel (excepto la categoría de control) del factor no es igual a la de la categoría de control. **< Pruebas de control** si la media en cualquier nivel del factor es menor que la de la categoría de control. **> Pruebas de control** si la media en cualquier nivel del factor es mayor que la de la categoría de control.

No asumiendo varianzas iguales

Las pruebas de comparación múltiple que no suponen varianzas iguales son las de Tamhane T2, Dunnett T3, Games-Howell y Dunnett *C*.

- *Tamhane's T2*. Prueba conservadora de comparaciones por parejas basada en una prueba t. Esta prueba es adecuada cuando las varianzas son desiguales.
- *Dunnett's T3*. Prueba de comparación por parejas basada en el módulo máximo estudentizado. Esta prueba es adecuada cuando las varianzas son desiguales.

- *Games-Howell*. Prueba de comparación por parejas que a veces es liberal. Esta prueba es adecuada cuando las varianzas son desiguales.
- *Dunnett's C*. Prueba de comparación por parejas basada en el rango estudentizado. Esta prueba es adecuada cuando las varianzas son desiguales.

Nota: Posiblemente le resulta más fácil interpretar el resultado de pruebas post hoc si deselecciona **Ocultar filas y columnas vacías** en el recuadro de diálogo Propiedades de tabla (en una tabla dinámica activada, elija **Propiedades de tabla** en el menú Formato).

Prueba de hipótesis nula

Especifica cómo se maneja el nivel de significación (alfa) para la prueba post hoc.

Utilizar el mismo nivel de significación (alfa) que el valor en Opciones
Cuando se selecciona esta opción, utiliza el mismo valor que se especifica en el diálogo Opciones.

Especificar el nivel de significación (alfa) para la prueba ad hoc
Cuando se selecciona, puede especificar el nivel de significación (alfa) en el campo **Nivel**.

Para obtener pruebas post hoc para ANOVA de un factor

ANOVA de un factor: Opciones

Estadísticos
Elija uno o más entre los siguientes:

Descriptivos
Calcula los siguientes estadísticos: Número de casos, Media, Desviación estándar, Error estándar de la media, Mínimo, Máximo y los Intervalos de confianza al 95% de cada variable dependiente para cada grupo.

Efectos fijos y aleatorios
Muestra la desviación estándar, el error estándar y un intervalo de confianza del 95% para el modelo de efectos fijos, y el error estándar, un intervalo de confianza del 95% y una estimación de la varianza entre componentes para el modelo de efectos aleatorios.

Prueba de homogeneidad de las varianzas
Calcula el estadístico de Levene para contrastar la igualdad de las varianzas de grupo. Esta prueba no depende del supuesto de normalidad.

Brown-Forsythe
Calcula el estadístico Brown-Forsythe para contrastar la igualdad de las medias de grupo. Este estadístico es preferible al estadístico *F* si no se supone la igualdad de las varianzas.

Welch
Calcula el estadístico Welch para probar la igualdad de las medias de grupo. Este estadístico es preferible al estadístico *F* si no se supone la igualdad de las varianzas.

Valores perdidos
Controla el tratamiento de los valores perdidos.

Excluir casos según análisis
Un caso con un valor perdido para la variable dependiente o bien la variable de factor para un análisis determinado no se utiliza en ese análisis. Además, los casos fuera del rango especificado para la variable de factor no se utilizan.

Excluir casos según lista
Los casos con valores perdidos para la variable de factor o para cualquier variable dependiente incluidos en la lista dependiente en el recuadro de diálogo principal se excluyen de todos los análisis. Si no se han especificado varias variables dependientes, esta opción no surte efecto.

Intervalo de confianza
De forma predeterminada, se muestra un intervalo de confianza al 95% para la diferencia entre la media y el valor de contraste hipotetizado. Introduzca un valor entre 1 y 99 para solicitar otro nivel de confianza.

Gráfico de las medias
Muestra un gráfico que representa las medias de subgrupo (las medias para cada grupo definido por valores de la variable de factor).

Para especificar las opciones de ANOVA de un factor

Características adicionales del comando ONEWAY

La sintaxis de comandos también le permite:

- Obtener estadísticos de efectos fijos y aleatorios. Desviación estándar, error estándar de la media e intervalos de confianza al 95% para el modelo de efectos fijos. Error estándar, intervalos de confianza al 95% y estimación de la varianza entre componentes para el modelo de efectos aleatorios (mediante `STATISTICS=EFFECTS`).
- Especificar niveles alfa para la menor diferencia de significación, Bonferroni, pruebas de comparación múltiple de Duncan y Scheffé (con el subcomando `RANGES`).
- Escribir una matriz de medias, desviaciones estándar y frecuencias, o leer una matriz de medias, frecuencias, varianzas combinadas y grados de libertad para las varianzas combinadas. Estas matrices pueden utilizarse en lugar de los datos en bruto para obtener un análisis de varianza de un factor (con el subcomando `MATRIX`).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

GLM Análisis univariante

El procedimiento GLM Univariante proporciona un análisis de regresión y un análisis de varianza para una variable dependiente mediante uno o más factores o variables. Las variables de factor dividen la población en grupos. Con el procedimiento Modelo lineal general se pueden contrastar hipótesis nulas sobre los efectos de otras variables en las medias de varias agrupaciones de una única variable dependiente. Se pueden investigar las interacciones entre los factores así como los efectos de los factores individuales, algunos de los cuales pueden ser aleatorios. Además, se pueden incluir los efectos de las covariables y las interacciones de covariables con los factores. Para el análisis de regresión, las variables (predictoras) independientes se especifican como covariables.

Se pueden contrastar tanto los modelos equilibrados como los no equilibrados. Se considera que un diseño está equilibrado si cada casilla del modelo contiene el mismo número de casos. Además de contrastar hipótesis, GLM Univariante genera estimaciones de los parámetros.

También se encuentran disponibles los contrastes a priori de uso más habitual para contrastar las hipótesis. Además, si una prueba *F* global ha mostrado cierta significación, pueden emplearse las pruebas post hoc para evaluar las diferencias entre las medias específicas. Las medias marginales estimadas ofrecen estimaciones de valores de las medias pronosticados para las casillas del modelo; los gráficos de perfil (gráficos de interacciones) de estas medias permiten observar fácilmente algunas de estas relaciones.

En su archivo de datos puede guardar residuos, valores pronosticados, distancia de Cook y valores de influencia como variables nuevas para comprobar los supuestos.

Ponderación MCP permite especificar una variable usada para aplicar a las observaciones una ponderación diferente en un análisis de mínimos cuadrados ponderados (MCP), por ejemplo para compensar la distinta precisión de las mediciones.

Ejemplo. Se recogen datos de los corredores individuales en el maratón de Chicago durante varios años. El tiempo final de cada corredor es la variable dependiente. Influyen otros factores como el clima (frío, calor o temperatura agradable), los meses de entrenamiento, el número de maratones anteriores y el

sexo. La edad se considera una covariable. Observará que el sexo es un efecto significativo y que la interacción del sexo con el clima es significativa.

Métodos. Las sumas de cuadrados de Tipo I, Tipo II, Tipo III y Tipo IV pueden emplearse para evaluar las diferentes hipótesis. Tipo III es el valor predeterminado.

Estadísticos. Las pruebas de rango post hoc y las comparaciones múltiples: Diferencia menos significativa (DMS), Bonferroni, Sidak, Scheffé, Múltiples *F* de Ryan-Einot-Gabriel-Welsch (R-E-G-W-F), Rango múltiple de Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls (S-N-K), Diferencia honestamente significativa de Tukey, *b* de Tukey, Duncan, GT2 de Hochberg, Gabriel, Pruebas *t* de Waller Duncan, Dunnett (unilateral y bilateral), T2 de Tamhane, T3 de Dunnett, Games-Howell y *C* de Dunnett. Estadísticos descriptivos: medias observadas, desviaciones estándar y frecuencias de todas las variables dependientes en todas las casillas. Las pruebas de Levene para la homogeneidad de varianzas.

Diagramas. Diagramas de dispersión por nivel, gráficos de residuos, gráficos de perfil (interacción).

GLM Univariante: Consideraciones sobre los datos

Datos. La variable dependiente es cuantitativa. Los factores son categóricos; pueden tener valores numéricos o valores de cadena de hasta ocho caracteres. Pueden tener valores numéricos o valores de cadena de hasta ocho caracteres. Las covariables son variables cuantitativas que están relacionadas con la variable dependiente.

Supuestos. Los datos son una muestra aleatoria de una población normal; en la población, todas las varianzas de las casillas son iguales. El análisis de varianza es robusto a las desviaciones de la normalidad, aunque los datos deberán ser simétricos. Para comprobar los supuestos, puede utilizar la prueba de homogeneidad de varianzas y los gráficos de dispersión por nivel. También puede examinar los residuos y los gráficos de residuos.

Para obtener un análisis GLM Univariante

1. En los menús seleccione:

   Analizar > **Modelo lineal general** > **Univariante...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una variable dependiente.
3. Seleccione variables para Factores fijos, Factores aleatorios y Covariables, en función de los datos.
4. Si lo desea, puede utilizar la Ponderación MCP para especificar una variable de ponderación para el análisis de mínimos cuadrados ponderados. Si el valor de la variable de ponderación es cero, negativo o perdido, el caso queda excluido del análisis. Una variable que ya se haya utilizado en el modelo no puede usarse como variable de ponderación.

GLM:Modelo

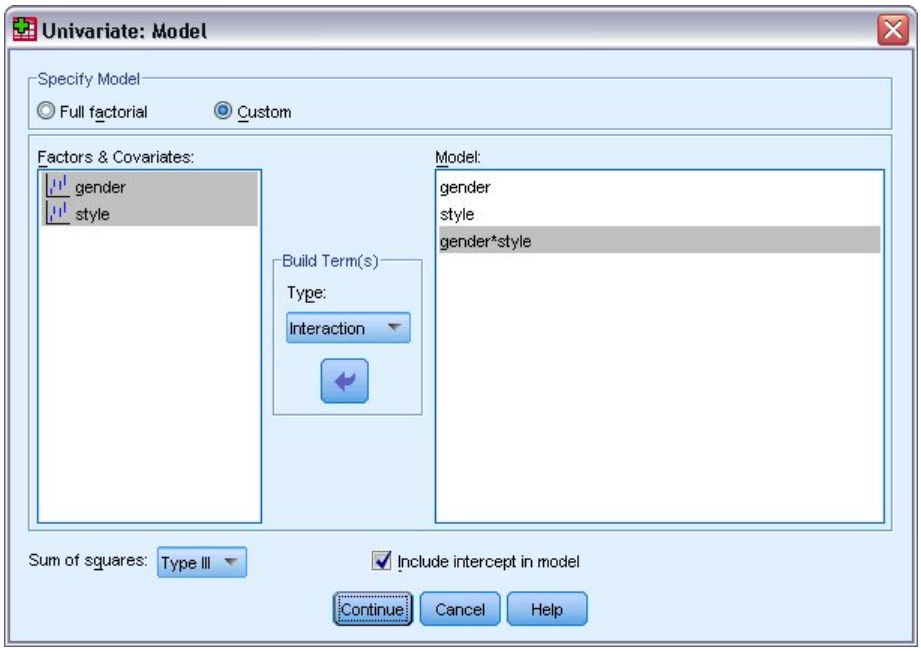

Figura 1. Cuadro de diálogo Univariante: Modelo

Especificar modelo. Un modelo factorial completo contiene todos los efectos principales del factor, todos los efectos principales de las covariables y todas las interacciones factor por factor. No contiene interacciones de covariable. Seleccione **Personalizado** para especificar sólo un subconjunto de interacciones o para especificar interacciones factor por covariable. Indique todos los términos que desee incluir en el modelo.

Factores y Covariables. Muestra una lista de los factores y las covariables.

Modelo. El modelo depende de la naturaleza de los datos. Después de seleccionar **Personalizado**, puede elegir los efectos principales y las interacciones que sean de interés para el análisis.

Suma de cuadrados Determina el método para calcular las sumas de cuadrados. Para los modelos equilibrados y no equilibrados sin casillas perdidas, el método de suma de cuadrados más utilizado es el de Tipo III.

Incluir la intersección en el modelo. La intersección se incluye normalmente en el modelo. Si supone que los datos pasan por el origen, puede excluir la intersección.

Construcción de términos y Términos personalizados

Crear términos
: Utilice esta opción si desea incluir términos no anidados de un tipo determinado (por ejemplo, efectos principales) para todas las combinaciones de un conjunto seleccionado de factores y covariables.

Construir términos personalizados
: Utilice esta opción si desea incluir términos anidados o si desea crear explícitamente una variable de término por variable. La creación de un término anidado implica los pasos siguientes:

Suma de cuadrados

Para el modelo, puede elegir un tipo de suma de cuadrados. El Tipo III es el más utilizado y es el tipo predeterminado.

Tipo I. Este método también se conoce como el método de descomposición jerárquica de la suma de cuadrados. Cada término se corrige sólo respecto al término que le precede en el modelo. El método Tipo I para la obtención de sumas de cuadrados se utiliza normalmente para:

- Un modelo ANOVA equilibrado en el que se especifica cualquier efecto principal antes de cualquier efecto de interacción de primer orden, cualquier efecto de interacción de primer orden se especifica antes de cualquier efecto de interacción de segundo orden, y así sucesivamente.
- Un modelo de regresión polinómica en el que se especifica cualquier término de orden inferior antes que cualquier término de orden superior.
- Un modelo puramente anidado en el que el primer efecto especificado está anidado dentro del segundo efecto especificado, el segundo efecto especificado está anidado dentro del tercero, y así sucesivamente. Esta forma de anidamiento solamente puede especificarse utilizando la sintaxis.

Tipo II. Este método calcula cada suma de cuadrados del modelo considerando sólo los efectos pertinentes. Un efecto pertinente es el que corresponde a todos los efectos que no contienen el que se está examinando. El método de suma de cuadrados de Tipo II se utiliza normalmente para:

- Un modelo ANOVA equilibrado.
- Cualquier modelo que sólo tenga efectos de factor principal.
- Cualquier modelo de regresión.
- Un diseño puramente anidado (esta forma de anidamiento solamente puede especificarse utilizando la sintaxis).

Tipo III. Es el método predeterminado. Este método calcula las sumas de cuadrados de un efecto de diseño como las sumas de cuadrados, corregidas respecto a cualquier otro efecto que no contenga el efecto, y ortogonales a cualquier efecto (si existe) que contenga el efecto. Las sumas de cuadrados de Tipo III tienen una gran ventaja por ser invariables respecto a las frecuencias de casilla, siempre que la forma general de estimabilidad permanezca constante. Así, este tipo de sumas de cuadrados se suele considerar de gran utilidad para un modelo no equilibrado sin casillas perdidas. En un diseño factorial sin casillas perdidas, este método equivale a la técnica de cuadrados ponderados de las medias de Yates. El método de suma de cuadrados de Tipo III se utiliza normalmente para:

- Cualquiera de los modelos que aparecen en los tipos I y II.
- Cualquier modelo equilibrado o desequilibrado sin casillas vacías.

Tipo IV. Este método está diseñado para una situación en la que hay casillas perdidas. Para cualquier efecto *F* en el diseño, si *F* no está contenida en cualquier otro efecto, entonces Tipo IV = Tipo III = Tipo II. Cuando *F* está contenida en otros efectos, el Tipo IV distribuye equitativamente los contrastes que se realizan entre los parámetros en *F* a todos los efectos de nivel superior. El método de suma de cuadrados de Tipo IV se utiliza normalmente para:

- Cualquiera de los modelos que aparecen en los tipos I y II.
- Cualquier modelo equilibrado o no equilibrado con casillas vacías.

GLM:Contrastes

Los contrastes se utilizan para contrastar las diferencias entre los niveles de un factor. Puede especificar un contraste para cada factor en el modelo (en un modelo de medidas repetidas, para cada factor inter-sujetos). Los contrastes representan las combinaciones lineales de los parámetros.

GLM Univariante. El contraste de hipótesis se basa en la hipótesis nula **LB** = 0, donde **L** es la matriz de coeficientes de contraste y **B** es el vector de parámetros. Cuando se especifica un contraste, se crea una matriz **L**. Las columnas de la matriz **L** correspondientes al factor coinciden con el contraste. El resto de las columnas se corrigen para que la matriz **L** sea estimable.

Los resultados incluyen un estadístico *F* para cada conjunto de contrastes. Para el contraste de diferencias también se muestran los intervalos de confianza simultáneos de tipo Bonferroni basados en la distribución *t* de Student.

Contrastes disponibles

Los contrastes disponibles son de desviación, simples, de diferencias, de Helmert, repetidos y polinómicos. En los contrastes de desviación y los contrastes simples, es posible determinar que la categoría de referencia sea la primera o la última categoría.

Tipos de contrastes

Desviación. Compara la media de cada nivel (excepto una categoría de referencia) con la media de todos los niveles (media global). Los niveles del factor pueden colocarse en cualquier orden.

Simples. Compara la media de cada nivel con la media de un nivel especificado. Este tipo de contraste resulta útil cuando existe un grupo de control. Puede seleccionar la primera o la última categoría como referencia.

Diferencia. Compara la media de cada nivel (excepto el primero) con la media de los niveles anteriores (a veces también se denominan contrastes de Helmert inversos).

Helmert. Compara la media de cada nivel del factor (excepto el último) con la media de los niveles siguientes.

Repetidas. Compara la media de cada nivel (excepto el último) con la media del nivel siguiente.

Polinómico. Compara el efecto lineal, cuadrático, cúbico, etc. El primer grado de libertad contiene el efecto lineal a través de todas las categorías; el segundo grado de libertad, el efecto cuadrático, y así sucesivamente. Estos contrastes se utilizan a menudo para estimar las tendencias polinómicas.

Gráficos de perfil GLM

Los gráficos de perfil (gráficos de interacción) sirven para comparar las medias marginales en el modelo. Un gráfico de perfil es un gráfico de líneas en el que cada punto indica la media marginal estimada de una variable dependiente (corregida respecto a las covariables) en un nivel de un factor. Los niveles de un segundo factor se pueden utilizar para generar líneas diferentes. Cada nivel en un tercer factor se puede utilizar para crear un gráfico diferente. Todos los factores fijos y aleatorios, si existen, están disponibles para los gráficos. Para los análisis multivariantes, los gráficos de perfil se crean para cada variable dependiente. En un análisis de medidas repetidas, es posible utilizar tanto los factores inter-sujetos como los inter-sujetos en los gráficos de perfil. Las opciones MLG - Multivariante y MLG - Medidas repetidas sólo estarán disponibles si tiene instalada la opción Estadísticas avanzadas.

Un gráfico de perfil de un factor muestra si las medias marginales estimadas aumentan o disminuyen a través de los niveles. Para dos o más factores, las líneas paralelas indican que no existe interacción entre los factores, lo que significa que puede investigar los niveles de un único factor. Las líneas no paralelas indican una interacción.

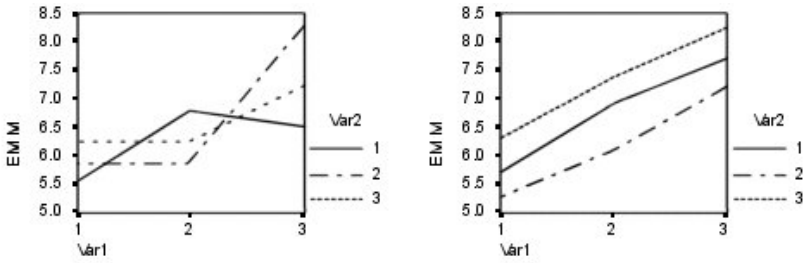

Figura 2. Gráfico no paralelo (izquierda) y gráfico paralelo (derecha)

Después de especificar un gráfico mediante la selección de los factores del eje horizontal y, de manera opcional, los factores para distintas líneas y gráficos, el gráfico deberá añadirse a la lista de gráficos.

Tipo de gráfico
El gráfico puede ser un gráfico de líneas o un gráfico de barras

Barras de error
Puede incluir barras de error que representen el intervalo de confianza o un número de errores estándar. El intervalo de confianza se basa en el nivel de significación especificado en el diálogo Opciones.

Incluir línea de referencia para media global
Incluye una línea de referencia que representa la media global general.

El eje Y empieza en 0

Para los gráficos de líneas con todos los valores positivos o todos los valores negativos, fuerza que el eje Y empiece en 0. Los gráficos de barras siempre empiezan en (o incluyen) 0.

Opciones MLG

Este cuadro de diálogo contiene estadísticos opcionales. Los estadísticos se calculan utilizando un modelo de efectos fijos.

Mostrar. Seleccione **Estadísticos descriptivos** para generar medias observadas, desviaciones estándar y frecuencias para cada variable dependiente en todas las casillas. La opción **Estimaciones del tamaño del efecto** ofrece un valor parcial de eta-cuadrado para cada efecto y cada estimación de parámetros. El estadístico eta cuadrado describe la proporción de variabilidad total atribuible a un factor. Seleccione **Potencia observada** para obtener la potencia de la prueba cuando la hipótesis alternativa se ha establecido basándose en el valor observado. Seleccione **Estimaciones de los parámetros** para generar las estimaciones de los parámetros, los errores estándar, las pruebas *t*, los intervalos de confianza y la potencia observada para cada prueba. Seleccione **Matriz de coeficientes de contraste** para obtener la matriz **L**.

Las **pruebas de homogeneidad** producen las pruebas de homogeneidad de varianzas de Levene para cada variable dependiente en todas las combinaciones de nivel de los factores inter-sujetos sólo para factores inter-sujetos. Las opciones de diagramas de dispersión por nivel y gráfico de los residuos son útiles para comprobar los supuestos sobre los datos. Estos elementos no estarán activado si no hay factores. Seleccione **Gráficos de los residuos** para generar un gráfico de los residuos observados respecto a los pronosticados respecto a los tipificados para cada variable dependiente. Estos gráficos son útiles para investigar el supuesto de varianzas iguales. Seleccione **Falta de ajuste** para comprobar si el modelo puede describir de forma adecuada la relación entre la variable dependiente y las variables independientes. Las **funciones estimables generales** permite construir pruebas de hipótesis personales basadas en las funciones estimables generales. Las filas en las matrices de coeficientes de contraste son combinaciones lineales de la función o funciones estimables generales.

Las **pruebas de heterocedasticidad** están disponibles para probar si la varianza de los errores (de cada variable dependiente) depende de los valores de las variables independientes. En las opciones **Test de Breusch-Pagan**, **Test de Breusch-Pagan modificada** y **Test F**, puede especificar el modelo en el que se base la prueba. De forma predeterminada, el modelo consta de un término constante, un término que es lineal en los valores pronosticados, un término que es cuadrático en los valores pronosticados, y un término de error.

En **Estimaciones de los parámetros con errores estándar robustos** se muestra una tabla de estimaciones de los parámetros, junto con errores estándar robustas o coherentes con la heterocedasticidad (HC); y los estadísticos *t*, valores de significación e intervalos de confianza que utilizan los errores estándar robustos. Hay disponibles cinco métodos distintos para poder realiza la estimación de matriz de covarianzas robustas.

HC0

En base a estimador asintótico original o robusto de gran muestra, empírico o "sandwich" de la matriz de covarianzas de las estimaciones de los parámetros. La parte central del sandwich contiene residuos de mínimos cuadrados ordinarios (OLS) o de mínimos cuadrados ponderados (WLS).

HC1

Una modificación de muestra finita de HC0, multiplicándola por N/(N-p), donde N es el tamaño de la muestra y p es el número de parámetros no redundantes en el modelo.

HC2

Una modificación de HC0 que conlleva dividir el cuadrado de los residuos por 1-h, donde h es la influencia del caso.

HC3

Una modificación de HC0 que se aproxima un estimador Jackknife. Los residuos al cuadrado se dividen por el cuadrado de 1-h.

HC4
Una modificación de HC0 que divide los residuos al cuadrado por 1-h a una potencia que varía según h, N y p, con un límite superior de 4.

Nivel de significación. Puede que le interese corregir el nivel de significación usado en las pruebas post hoc y el nivel de confianza empleado para construir intervalos de confianza. El valor especificado también se utiliza para calcular la potencia observada para la prueba. Si especifica un nivel de significación, el cuadro de diálogo mostrará el nivel asociado de los intervalos de confianza.

Características adicionales del comando UNIANOVA

La sintaxis de comandos también le permite:

- Especificar efectos anidados en el diseño (utilizando el subcomando DESIGN).
- Especificar contrastes de los efectos respecto a una combinación lineal de efectos o un valor (utilizando el subcomando TEST).
- Especificar varios contrastes (utilizando el subcomando CONTRAST).
- Incluir los valores perdidos del usuario (utilizando el subcomando MISSING).
- Especificar criterios EPS (mediante el subcomando CRITERIA).
- Construir una matriz **L**, **M** o **K** personalizada (utilizando los subcomandos LMATRIX, MMATRIX y KMATRIX).
- Para contrastes de desviación o simples, especifique una categoría de referencia intermedia (utilizando el subcomando CONTRAST).
- Especificar métricas para contrastes polinómicos (utilizando el subcomando CONTRAST).
- Especificar términos de error para comparaciones post hoc (utilizando el subcomando POSTHOC).
- Calcular medias marginales estimadas para cualquier factor o interacción de factores entre los factores de la lista de factores (utilizando el subcomando EMMEANS).
- Especificar nombres para variables temporales (utilizando el subcomando SAVE).
- Construir un archivo de datos de matriz de correlaciones (utilizando el subcomando OUTFILE).
- Construir un archivo de datos de matriz que contenga estadísticos de la tabla de ANOVA inter-sujetos (utilizando el subcomando OUTFILE).
- Guardar la matriz del diseño en un nuevo archivo de datos (utilizando el subcomando OUTFILE).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

GLM:Comparaciones post hoc

Pruebas de comparaciones múltiples post hoc Una vez que se ha determinado que existen diferencias entre las medias, las pruebas de rango post hoc y las comparaciones múltiples por parejas permiten determinar qué medias difieren. Las comparaciones se realizan sobre valores sin corregir. Estas pruebas se utilizan únicamente para factores inter-sujetos fijos. En GLM Medidas repetidas, estas pruebas no están disponibles si no existen factores inter-sujetos y las pruebas de comparación múltiple post hoc se realizan para la media a través de los niveles de los factores intra-sujetos. Para GLM - Multivariante, las pruebas post hoc se realizan por separado para cada variable dependiente. Las opciones MLG - Multivariante y MLG - Medidas repetidas sólo estarán disponibles si tiene instalada la opción Estadísticas avanzadas.

Las pruebas de diferencia honestamente significativa de Tukey y de Bonferroni son pruebas de comparación múltiple muy utilizadas. La **prueba de Bonferroni**, basada en el estadístico *t* de Student, corrige el nivel de significación observado por el hecho de que se realizan comparaciones múltiples. La **prueba t de Sidak** también corrige el nivel de significación y da lugar a límites más estrechos que los de Bonferroni. La **prueba de diferencia honestamente significativa de Tukey** utiliza el estadístico del rango estudentizado para realizar todas las comparaciones por pares entre los grupos y establece la tasa de error por experimento como la tasa de error para el conjunto de todas las comparaciones por pares. Cuando se contrasta un gran número de pares de medias, la prueba de la diferencia honestamente

significativa de Tukey es más potente que la prueba de Bonferroni. Para un número reducido de pares, Bonferroni es más potente.

GT2 de Hochberg es similar a la prueba de la diferencia honestamente significativa de Tukey, pero se utiliza el módulo máximo estudentizado. La prueba de Tukey suele ser más potente. La **prueba de comparación por parejas de Gabriel** también utiliza el módulo máximo estudentizado y es generalmente más potente que la GT2 de Hochberg cuando los tamaños de las casillas son desiguales. La prueba de Gabriel se puede convertir en liberal cuando los tamaños de las casillas varían mucho.

La **prueba t de comparación múltiple por parejas de Dunnett** compara un conjunto de tratamientos con una media de control simple. La última categoría es la categoría de control predeterminada. Si lo desea, puede seleccionar la primera categoría. Asimismo, puede elegir una prueba unilateral o bilateral. Para comprobar que la media de cualquier nivel del factor (excepto la categoría de control) no es igual a la de la categoría de control, utilice una prueba bilateral. Para probar si la media en cualquier nivel del factor es menor que la de la categoría de control, seleccione **< Control**. Asimismo, para probar si la media en cualquier nivel del factor es mayor que la de la categoría de control, seleccione **> Control**.

Ryan, Einot, Gabriel y Welsch (R-E-G-W) desarrollaron dos pruebas de rangos múltiples por pasos. Los procedimientos múltiples por pasos (por tamaño de las distancias) contrastan en primer lugar si todas las medias son iguales. Si no son iguales, se contrasta la igualdad en los subconjuntos de medias. **R-E-G-W F** se basa en una prueba *F* y **R-E-G-W Q** se basa en un rango estudentizado. Estas pruebas son más potentes que la prueba de rangos múltiples de Duncan y Student-Newman-Keuls (que también son procedimientos múltiples por pasos), pero no se recomiendan para tamaños de casillas desiguales.

Cuando las varianzas son desiguales, utilice **T2 de Tamhane** (prueba conservadora de comparación por parejas basada en una prueba *t*), **T3 de Dunnett** (prueba de comparación por parejas basada en el módulo máximo estudentizado), **prueba de comparación por parejas Games-Howell** (a veces liberal), o **C de Dunnett** (prueba de comparación por parejas basada en el rango estudentizado). Tenga en cuenta que estas pruebas no son válidas y no se realizarán si el modelo tiene múltiples factores.

La **prueba de rango múltiple de Duncan**, Student-Newman-Keuls (**S-N-K**) y **b de Tukey** son pruebas de rango que asignan rangos a medias de grupo y calculan un valor de rango. Estas pruebas no se utilizan con la misma frecuencia que las pruebas anteriormente mencionadas.

La **prueba t de Waller-Duncan** utiliza la aproximación bayesiana. Esta prueba de rango emplea la media armónica del tamaño de la muestra cuando los tamaños muestrales no son iguales.

El nivel de significación de la prueba de **Scheffé** está diseñado para permitir todas las combinaciones lineales posibles de las medias de grupo que se van a contrastar, no sólo las comparaciones por parejas disponibles en esta característica. El resultado es que la prueba de Scheffé es normalmente más conservadora que otras pruebas, lo que significa que se precisa una mayor diferencia entre las medias para la significación.

La prueba de comparación múltiple por parejas de la diferencia menos significativa (**DMS**) es equivalente a varias pruebas *t* individuales entre todos los pares de grupos. La desventaja de esta prueba es que no se realiza ningún intento de corregir el nivel de significación observado para realizar las comparaciones múltiples.

Pruebas mostradas. Se proporcionan comparaciones por parejas para DMS, Sidak, Bonferroni, Games-Howell, T2 y T3 de Tamhane, *C* de Dunnett y T3 de Dunnett. También se facilitan subconjuntos homogéneos para S-N-K, *b* de Tukey, Duncan, R-E-G-W *F*, R-E-G-W *Q* y Waller. La prueba de la diferencia honestamente significativa de Tukey, la GT2 de Hochberg, la prueba de Gabriel y la prueba de Scheffé son pruebas de comparaciones múltiples y pruebas de rango.

Opciones MLG

Este cuadro de diálogo contiene estadísticos opcionales. Los estadísticos se calculan utilizando un modelo de efectos fijos.

Mostrar. Seleccione **Estadísticos descriptivos** para generar medias observadas, desviaciones estándar y frecuencias para cada variable dependiente en todas las casillas. La opción **Estimaciones del tamaño del efecto** ofrece un valor parcial de eta-cuadrado para cada efecto y cada estimación de parámetros. El estadístico eta cuadrado describe la proporción de variabilidad total atribuible a un factor. Seleccione

Potencia observada para obtener la potencia de la prueba cuando la hipótesis alternativa se ha establecido basándose en el valor observado. Seleccione **Estimaciones de los parámetros** para generar las estimaciones de los parámetros, los errores estándar, las pruebas *t*, los intervalos de confianza y la potencia observada para cada prueba. Seleccione **Matriz de coeficientes de contraste** para obtener la matriz **L**.

Las **pruebas de homogeneidad** producen las pruebas de homogeneidad de varianzas de Levene para cada variable dependiente en todas las combinaciones de nivel de los factores inter-sujetos sólo para factores inter-sujetos. Las opciones de diagramas de dispersión por nivel y gráfico de los residuos son útiles para comprobar los supuestos sobre los datos. Estos elementos no estarán activado si no hay factores. Seleccione **Gráficos de los residuos** para generar un gráfico de los residuos observados respecto a los pronosticados respecto a los tipificados para cada variable dependiente. Estos gráficos son útiles para investigar el supuesto de varianzas iguales. Seleccione **Falta de ajuste** para comprobar si el modelo puede describir de forma adecuada la relación entre la variable dependiente y las variables independientes. Las **funciones estimables generales** permite construir pruebas de hipótesis personales basadas en las funciones estimables generales. Las filas en las matrices de coeficientes de contraste son combinaciones lineales de la función o funciones estimables generales.

Las **pruebas de heterocedasticidad** están disponibles para probar si la varianza de los errores (de cada variable dependiente) depende de los valores de las variables independientes. En las opciones **Test de Breusch-Pagan**, **Test de Breusch-Pagan modificada** y **Test F**, puede especificar el modelo en el que se base la prueba. De forma predeterminada, el modelo consta de un término constante, un término que es lineal en los valores pronosticados, un término que es cuadrático en los valores pronosticados, y un término de error.

En **Estimaciones de los parámetros con errores estándar robustos** se muestra una tabla de estimaciones de los parámetros, junto con errores estándar robustas o coherentes con la heterocedasticidad (HC); y los estadísticos *t*, valores de significación e intervalos de confianza que utilizan los errores estándar robustos. Hay disponibles cinco métodos distintos para poder realiza la estimación de matriz de covarianzas robustas.

HC0
: En base a estimador asintótico original o robusto de gran muestra, empírico o "sandwich" de la matriz de covarianzas de las estimaciones de los parámetros. La parte central del sandwich contiene residuos de mínimos cuadrados ordinarios (OLS) o de mínimos cuadrados ponderados (WLS).

HC1
: Una modificación de muestra finita de HC0, multiplicándola por N/(N-p), donde N es el tamaño de la muestra y p es el número de parámetros no redundantes en el modelo.

HC2
: Una modificación de HC0 que conlleva dividir el cuadrado de los residuos por 1-h, donde h es la influencia del caso.

HC3
: Una modificación de HC0 que se aproxima un estimador Jackknife. Los residuos al cuadrado se dividen por el cuadrado de 1-h.

HC4
: Una modificación de HC0 que divide los residuos al cuadrado por 1-h a una potencia que varía según h, N y p, con un límite superior de 4.

Nivel de significación. Puede que le interese corregir el nivel de significación usado en las pruebas post hoc y el nivel de confianza empleado para construir intervalos de confianza. El valor especificado también se utiliza para calcular la potencia observada para la prueba. Si especifica un nivel de significación, el cuadro de diálogo mostrará el nivel asociado de los intervalos de confianza.

Características adicionales del comando UNIANOVA

La sintaxis de comandos también le permite:

- Especificar efectos anidados en el diseño (utilizando el subcomando DESIGN).

- Especificar contrastes de los efectos respecto a una combinación lineal de efectos o un valor (utilizando el subcomando TEST).
- Especificar varios contrastes (utilizando el subcomando CONTRAST).
- Incluir los valores perdidos del usuario (utilizando el subcomando MISSING).
- Especificar criterios EPS (mediante el subcomando CRITERIA).
- Construir una matriz **L**, **M** o **K** personalizada (utilizando los subcomandos LMATRIX, MMATRIX y KMATRIX).
- Para contrastes de desviación o simples, especifique una categoría de referencia intermedia (utilizando el subcomando CONTRAST).
- Especificar métricas para contrastes polinómicos (utilizando el subcomando CONTRAST).
- Especificar términos de error para comparaciones post hoc (utilizando el subcomando POSTHOC).
- Calcular medias marginales estimadas para cualquier factor o interacción de factores entre los factores de la lista de factores (utilizando el subcomando EMMEANS).
- Especificar nombres para variables temporales (utilizando el subcomando SAVE).
- Construir un archivo de datos de matriz de correlaciones (utilizando el subcomando OUTFILE).
- Construir un archivo de datos de matriz que contenga estadísticos de la tabla de ANOVA inter-sujetos (utilizando el subcomando OUTFILE).
- Guardar la matriz del diseño en un nuevo archivo de datos (utilizando el subcomando OUTFILE).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

MLG: Guardar

Es posible guardar los valores pronosticados por el modelo, los residuos y las medidas relacionadas como variables nuevas en el Editor de datos. Muchas de estas variables se pueden utilizar para examinar supuestos sobre los datos. Si desea almacenar los valores para utilizarlos en otra sesión de IBM SPSS Estadística , guárdelos en el archivo de datos actual.

Valores pronosticados. Son los valores que predice el modelo para cada caso.

- *Sin estandarizar*. Valor predicho por el modelo para la variable dependiente.
- *Ponderado*. Valores pronosticados no tipificados ponderados. Sólo están disponibles si se seleccionó previamente una variable de ponderación MCP.
- *Error estándar*. Estimación de la desviación estándar del valor promedio de la variable dependiente para los casos que tengan los mismos valores en las variables independientes.

Diagnósticos. Son medidas para identificar casos con combinaciones poco usuales de valores para los casos y las variables independientes que puedan tener un gran impacto en el modelo.

- *Distancia de Cook*. Una medida de cuánto cambiarían los residuos de todos los casos si se excluyera un caso concreto del cálculo de los coeficientes de regresión. Una Distancia de Cook grande indica que la exclusión de ese caso del cálculo de los estadísticos de regresión hará variar substancialmente los coeficientes.
- *Valores de influencia*. Valores de influencia no centrados. La influencia relativa de una observación en el ajuste del modelo.

Residuos. Un residuo no tipificado es el valor real de la variable dependiente menos el valor predicho por el modelo. También se encuentran disponibles residuos eliminados, estudentizados y estandarizados. Si ha seleccionado una variable MCP, contará además con residuos no tipificados ponderados.

- *Sin estandarizar*. Diferencia entre un valor observado y el valor predicho por el modelo.
- *Ponderado*. Residuos no tipificados ponderados. Sólo están disponibles si se seleccionó previamente una variable de ponderación MCP.

- *Estandarizado*. El residuo dividido por una estimación de su desviación estándar. Los residuos estandarizados, que son conocidos también como los residuos de Pearson o residuos estandarizados, tienen una media de 0 y una desviación estándar de 1.
- *Estudentizado*. El residuo dividido por una estimación de su desviación estándar que varía de un caso a otro, dependiendo de la distancia de los valores de cada caso en las variables independientes de las medias de las variables independientes. A veces se hace referencia a ellos como residuos estudentizados internamente.
- *Suprimido*. El residuo de un caso cuando ese caso se excluye del cálculo de los coeficientes de regresión. Es igual a la diferencia entre el valor de la variable dependiente y el valor predicho corregido.

Estadísticos de los coeficientes. Escribe una matriz varianza-covarianza de las estimaciones de los parámetros del modelo en un nuevo conjunto de datos de la sesión actual o un archivo de datos externo de IBM SPSS Estadística . Asimismo, para cada variable dependiente habrá una fila de estimaciones de los parámetros, una fila de errores estándar de las estimaciones de los parámetros, una fila de valores de significación para los estadísticos *t* correspondientes a las estimaciones de los parámetros y una fila de grados de libertad de los residuos. En un modelo multivariante, existen filas similares para cada variable dependiente. Cuando se selecciona la estadística coherente de heterocedasticidad (sólo disponible para los modelos univariante), se calcula la matriz varianza-covarianza utilizando un estimador robusto, la fila de errores estándar muestra los errores estándar robustos, y los valores de significación reflejan los errores robustos. Si lo desea, puede usar este archivo matricial en otros procedimientos que lean archivos matriciales.

Medias marginales estimadas GLM

Seleccione los factores e interacciones para los que desee obtener estimaciones de las medias marginales de la población en las casillas. Estas medias se corrigen respecto a las covariables, si las hay.

Comparar los efectos principales
Proporciona comparaciones por parejas no corregidas entre las medias marginales estimadas para cualquier efecto principal del modelo, tanto para los factores inter-sujetos como para los intra-sujetos. Este elemento sólo se encuentra disponible si los efectos principales están seleccionados en la lista Mostrar las medias para.

Comparar efectos principales simples
El ajuste se activa siempre que la lista objetivo contenga uno o más efectos de producto o de interacción (por ejemplo, A*B, A*B*C). El ajuste admite la especificación de comparaciones entre efectos principales simples, que son efectos principales anidados dentro de los niveles de otros factores.

Ajuste de intervalo de confianza
Seleccione un ajuste de diferencia menor significativa (DMS), Bonferroni o Sidak para los intervalos de confianza y la significación. Este elemento solo estará disponible si se selecciona **Comparar los efectos principales** y/o **Comparar efectos principales simples**.

Especificación de medias marginales estimadas

1. En los menús, elija uno de los procedimientos disponibles en > **Analizar** > **Modelo lineal general**.
2. En el cuadro de diálogo principal, pulse **Medias ME**.

Opciones MLG

Este cuadro de diálogo contiene estadísticos opcionales. Los estadísticos se calculan utilizando un modelo de efectos fijos.

Mostrar. Seleccione **Estadísticos descriptivos** para generar medias observadas, desviaciones estándar y frecuencias para cada variable dependiente en todas las casillas. La opción **Estimaciones del tamaño del efecto** ofrece un valor parcial de eta-cuadrado para cada efecto y cada estimación de parámetros. El estadístico eta cuadrado describe la proporción de variabilidad total atribuible a un factor. Seleccione **Potencia observada** para obtener la potencia de la prueba cuando la hipótesis alternativa se ha

establecido basándose en el valor observado. Seleccione **Estimaciones de los parámetros** para generar las estimaciones de los parámetros, los errores estándar, las pruebas *t*, los intervalos de confianza y la potencia observada para cada prueba. Seleccione **Matriz de coeficientes de contraste** para obtener la matriz **L**.

Las **pruebas de homogeneidad** producen las pruebas de homogeneidad de varianzas de Levene para cada variable dependiente en todas las combinaciones de nivel de los factores inter-sujetos sólo para factores inter-sujetos. Las opciones de diagramas de dispersión por nivel y gráfico de los residuos son útiles para comprobar los supuestos sobre los datos. Estos elementos no estarán activado si no hay factores. Seleccione **Gráficos de los residuos** para generar un gráfico de los residuos observados respecto a los pronosticados respecto a los tipificados para cada variable dependiente. Estos gráficos son útiles para investigar el supuesto de varianzas iguales. Seleccione **Falta de ajuste** para comprobar si el modelo puede describir de forma adecuada la relación entre la variable dependiente y las variables independientes. Las **funciones estimables generales** permite construir pruebas de hipótesis personales basadas en las funciones estimables generales. Las filas en las matrices de coeficientes de contraste son combinaciones lineales de la función o funciones estimables generales.

Las **pruebas de heterocedasticidad** están disponibles para probar si la varianza de los errores (de cada variable dependiente) depende de los valores de las variables independientes. En las opciones **Test de Breusch-Pagan**, **Test de Breusch-Pagan modificada** y **Test F**, puede especificar el modelo en el que se base la prueba. De forma predeterminada, el modelo consta de un término constante, un término que es lineal en los valores pronosticados, un término que es cuadrático en los valores pronosticados, y un término de error.

En **Estimaciones de los parámetros con errores estándar robustos** se muestra una tabla de estimaciones de los parámetros, junto con errores estándar robustas o coherentes con la heterocedasticidad (HC); y los estadísticos *t*, valores de significación e intervalos de confianza que utilizan los errores estándar robustos. Hay disponibles cinco métodos distintos para poder realiza la estimación de matriz de covarianzas robustas.

HC0
: En base a estimador asintótico original o robusto de gran muestra, empírico o "sandwich" de la matriz de covarianzas de las estimaciones de los parámetros. La parte central del sandwich contiene residuos de mínimos cuadrados ordinarios (OLS) o de mínimos cuadrados ponderados (WLS).

HC1
: Una modificación de muestra finita de HC0, multiplicándola por N/(N-p), donde N es el tamaño de la muestra y p es el número de parámetros no redundantes en el modelo.

HC2
: Una modificación de HC0 que conlleva dividir el cuadrado de los residuos por 1-h, donde h es la influencia del caso.

HC3
: Una modificación de HC0 que se aproxima un estimador Jackknife. Los residuos al cuadrado se dividen por el cuadrado de 1-h.

HC4
: Una modificación de HC0 que divide los residuos al cuadrado por 1-h a una potencia que varía según h, N y p, con un límite superior de 4.

Nivel de significación. Puede que le interese corregir el nivel de significación usado en las pruebas post hoc y el nivel de confianza empleado para construir intervalos de confianza. El valor especificado también se utiliza para calcular la potencia observada para la prueba. Si especifica un nivel de significación, el cuadro de diálogo mostrará el nivel asociado de los intervalos de confianza.

Modelo de regresión auxiliar GLM

En el cuadro de diálogo Modelo de regresión auxiliar se especifica el modelo que se utiliza para probar la heterocedasticidad.

Utilizar valores pronosticados
Utiliza un modelo que consta de un término constante, un término que es lineal en los valores pronosticados, un término que es cuadrático en los valores pronosticados, y un término de error.

Utilizar modelo univariado
Utiliza el modelo que se especifica en el subdiálogo Modelo. Se incluye un término de intersección si el modelo especificado no contiene ninguno.

Modelo personalizado
Utiliza el modelo que se especifique explícitamente.

Crear términos
Utilice esta opción si desea incluir términos no anidados de un tipo determinado (por ejemplo, efectos principales) para todas las combinaciones de un conjunto seleccionado de factores y covariables.

Construir términos personalizados
Utilice esta opción si desea incluir términos anidados o si desea crear explícitamente una variable de término por variable. La creación de un término anidado implica los pasos siguientes:

Características adicionales del comando UNIANOVA

La sintaxis de comandos también le permite:

- Especificar efectos anidados en el diseño (utilizando el subcomando DESIGN).
- Especificar contrastes de los efectos respecto a una combinación lineal de efectos o un valor (utilizando el subcomando TEST).
- Especificar varios contrastes (utilizando el subcomando CONTRAST).
- Incluir los valores perdidos del usuario (utilizando el subcomando MISSING).
- Especificar criterios EPS (mediante el subcomando CRITERIA).
- Construir una matriz **L**, **M** o **K** personalizada (utilizando los subcomandos LMATRIX, MMATRIX y KMATRIX).
- Para contrastes de desviación o simples, especifique una categoría de referencia intermedia (utilizando el subcomando CONTRAST).
- Especificar métricas para contrastes polinómicos (utilizando el subcomando CONTRAST).
- Especificar términos de error para comparaciones post hoc (utilizando el subcomando POSTHOC).
- Calcular medias marginales estimadas para cualquier factor o interacción de factores entre los factores de la lista de factores (utilizando el subcomando EMMEANS).
- Especificar nombres para variables temporales (utilizando el subcomando SAVE).
- Construir un archivo de datos de matriz de correlaciones (utilizando el subcomando OUTFILE).
- Construir un archivo de datos de matriz que contenga estadísticos de la tabla de ANOVA inter-sujetos (utilizando el subcomando OUTFILE).
- Guardar la matriz del diseño en un nuevo archivo de datos (utilizando el subcomando OUTFILE).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Bland-Altman

El análisis de Bland Altman es una técnica gráfica para evaluar el sesgo entre las diferencias de medias. Altman y Bland propusieron este análisis en el año 1983. El análisis de Bland Altman se basa en la cuantificación del acuerdo entre dos mediciones cuantitativas mediante la exploración de las diferencias medias y la construcción de los límites de acuerdo.

Ejemplo

A menudo, los laboratorios médicos necesitan evaluar la concordancia entre dos métodos de medición, elegir entre dos métodos o evaluar un método nuevo o alternativo. Además, cuando hay un problema de

alineación entre dos instrumentos, se necesitan algunas herramientas para medir y valorar las diferencias y la causa de estas diferencias.

Este procedimiento requiere Statistics Base Edition.

En los menús elija: **AnalizarEstadística descriptiva** > **Análisis de Bland Altman**...

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

Consideraciones sobre los datos: Las variables están en numérico.

1. Selecciona las variables y añádelas a la **Variables emparejadas:**.
2. Si selecciona **OK**, el procedimiento se ejecuta teniendo en cuenta la configuración predeterminada.
3. Puede elegir entre **Aplicar transformación de registro** o **Sólo emparejado** o ambas.
4. Pulse **Aceptar**.

Aplique la transformación logarítmica: Se recomienda la **transformación logarítmica**, si las diferencias no se distribuyen normalmente.

Sólo pareado: Esta opción se especifica cuando la primera variable de la primera columna se compara con la primera variable de la segunda columna.

Nota: De las variables emparejadas, el análisis se realiza sólo hasta el último valor del conjunto de variables que contiene el menor número de valores. El análisis se realiza en su totalidad para conjuntos de variables emparejadas que contienen el mismo número de valores.

Bland Altman: Criterios

Puede cambiar la configuración predeterminada en la pestaña **Bland Atman: Criterios**.

1. Por defecto, la opción de mostrar los ratios en porcentaje no está seleccionada. Si quieres añadirlo, selecciona **Ratio como en porcentajes**.
2. Compruebe la configuración de **Ámbito** y **Límites del acuerdo**, y modifíquela si es necesario.
3. Pulse **Continuar**.

Bland Altman: Imprimir

1. Marque las opciones que desea incluir en la salida.
2. Si selecciona **Imprimir detalles de análisis**, la configuración predeterminada para el número máximo de filas que se mostrarán es 10. Cambie el número para tener más o menos filas en la pantalla. El máximo de filas de datos que se pueden mostrar es 1000. **Orden** puede ser **Basado en** las cuatro opciones de lista y en **Orden ascendente o descendente**.

   Por defecto **Ordenar** se establece en base a la primera medida y en orden descendente.
3. Pulse **Continuar**.

Bland Altman: Argumento

Puede cambiar la configuración predeterminada en la pestaña **Análisis Atman: Plot** tab.

1. Elija los ajustes deseados, incluyendo **Plots de razón**, la **Línea de referencia en la diferencia cero**, **Límites de acuerdo** e **Intervalos de confianza de las diferencias medias**.
2. Pulse **Continuar**.

Prueba de chi cuadrado

El procedimiento de la prueba ji cuadrado tabula una variable en categorías y calcula un estadístico ji cuadrado. La prueba chi-cuadrado es una prueba estadística de hipótesis utilizada para determinar si existe una asociación significativa entre dos variables categóricas. La prueba evalúa si las frecuencias

observadas en los datos difieren de las frecuencias esperadas bajo la hipótesis nula, que supone la ausencia de asociación entre las variables.

Este procedimiento requiere Statistics Base Edition.

Ejemplo
En los estudios de mercado, los analistas deben comprender las preferencias de los consumidores. Para ello, comparan el comportamiento observado de los clientes con los patrones esperados para crear estrategias de marketing eficaces.

Estadísticos
El chi-cuadrado de Pearson, el chi-cuadrado de la razón de verosimilitud, la prueba exacta de Fisher y el chi-cuadrado corregido de Yates (corrección de continuidad).

Consideraciones sobre los datos y supuestos
Use variables categóricas numéricas ordenadas o no ordenadas (niveles de medición ordinal o nominal). Para convertir las variables de cadena en variables numéricas, utilice el procedimiento Recodificación automática, disponible en el menú Transformar.

Las pruebas no paramétricas no requieren supuestos sobre la forma de la distribución subyacente. Se asume que los datos son una muestra aleatoria. Las frecuencias esperadas para cada categoría deben ser al menos 1. No más del 20 % de las categorías deberían tener frecuencias esperadas inferiores a 5.

En el menú, haga clic en **Analizar** > **estadísticas** > **descriptivasChiCuadrado**.

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

1. Selecciona las variables y añádelas a **Fila(s)** y **Columna(s)**.
2. Pulse **Aceptar**. El procedimiento se ejecuta teniendo en cuenta la configuración por defecto.

Prueba Chi Cuadrado: Exacto

- *Asintótica*. El nivel de significación basado en la distribución asintótica de una estadística de prueba. Normalmente un valor menor que 0,05 se considera significativo. La significación asintótica se basa en la asunción de que el conjunto de datos es grande. Si el conjunto de datos es pequeño o se distribuye de manera pobre, puede que no sea un buen indicador de la significación.
- *Estimación de Monte Carlo*. Estimación imparcial del nivel de significación exacto, calculada mediante muestreo repetido a partir de un conjunto de tablas de referencia con las mismas dimensiones y márgenes de fila y columna que la tabla observada. El método de Monte Carlo permite estimar la significación exacta sin tener que confiar en los supuestos que requiere el método asintótico. Este método es más útil si el conjunto de datos es demasiado grande para calcular la significación exacta, pero los datos no cumplen los supuestos del método asintótico.
- *Exacta*. Se calcula de forma exacta la probabilidad del resultado observado o un resultado más extremo. Típicamente, un valor de significación menor que 0,05 se considera significativo, indicando que hay alguna relación entre las variables de fila y de columna.

Prueba Chi Cuadrado: Células

El procedimiento Chi Cuadrado muestra las frecuencias esperadas. Cada celda de la tabla puede contener cualquier combinación de recuentos y porcentajes.

Recuentos
Número de casos realmente observados y número de casos esperados si las variables de fila y columna son independientes entre sí. Puede optar por ocultar recuentos inferiores un número entero especificado. Los valores ocultos se visualizan como **<N**, donde **N** es el entero especificado. El entero especificado debe ser mayor o igual a 2, aunque se permite el valor 0 y especifica que no se hay recuentos ocultos.

Porcentajes
Los porcentajes se pueden sumar a través de las filas o a lo largo de las columnas. También se encuentran disponibles los porcentajes del número total de casos representados en la tabla (una capa).

Nota: Si **Ocultar recuentos pequeños** está seleccionado en el grupo Recuentos, se ocultarán también los porcentajes asociados con recuentos ocultos.

Crear tabla de estilos APA
Crea tablas de salida que cumplen las directrices de estilo APA.

Nota: Las opciones **Observado**, **Esperado**, **Fila**, **Columna**y **Total** no están disponibles cuando se selecciona **Crear tabla de estilo APA**.

Prueba Chi Cuadrado: Formato

Con la opción de formato, puede especificar el orden de las variables que se visualizan en los resultados de la prueba.

Vaya a **Analizar** > **estadísticas descriptivas** > **ChiCuadrado**. Haga clic en **Formato** en el cuadro de diálogo **Prueba Chi Cuadrado**.

Puede ordenar las filas en orden ascendente o descendente de los valores de la variable de fila. En **Orden de filas**, seleccione **Ascendente** o **Descendente**.

Correlaciones bivariadas

El procedimiento Correlaciones bivariadas calcula el coeficiente de correlación de Pearson, la rho de Spearman y la tau-*b* de Kendall con sus niveles de significación. Las correlaciones miden cómo están relacionadas las variables o los órdenes de los rangos. Antes de calcular un coeficiente de correlación, inspeccione los datos para detectar valores atípicos (que pueden generar resultados equívocos) y evidencias de una relación lineal. El coeficiente de correlación de Pearson es una medida de asociación lineal. Dos variables pueden estar perfectamente relacionadas, pero si la relación no es lineal, el coeficiente de correlación de Pearson no será un estadístico adecuado para medir su asociación.

Los valores de intervalo de confianza están disponibles para Pearson y Spearman.

Ejemplo
¿Está el número de partidos ganados por un equipo de baloncesto correlacionado con el número medio de puntos anotados por partido? Un diagrama de dispersión indica que existe una relación lineal. Al analizar los datos de la temporada 1994–1995 de la NBA, se descubre que el coeficiente de correlación de Pearson (0,581) es significativo al nivel 0,01. Se puede sospechar que cuantos más partidos se ganen por temporada, menos puntos habrán anotado los adversarios. Estas variables están correlacionadas negativamente (-0,401) y la correlación es significativa al nivel 0,05.

Estadísticos
Para cada variable: número de casos sin valores perdidos, media y desviación estándar. Para cada par de variables: coeficiente de correlación de Pearson, rho de Spearman, tau-*b* de Kendall, productos vectoriales de las desviaciones y covarianzas.

Consideraciones sobre los datos

Datos
Utilice variables cuantitativas simétricas para el coeficiente de correlación de Pearson y variables cuantitativas o variables con categorías ordenadas para la rho de Spearman y la tau-*b* de Kendall.

Supuestos
El coeficiente de correlación de Pearson asume que cada pareja de variables es normal bivariante.

Obtención de correlaciones bivariadas

En los menús seleccione:

Analizar > **Correlaciones** > **Bivariadas...**

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

1. Seleccione dos o más variables numéricas.

   También se encuentran disponibles las siguientes opciones:

   Coeficientes de correlación
   Para variables cuantitativas distribuidas normalmente, elija el coeficiente de correlación de **Pearson**. Si los datos no están normalmente distribuidos o tienen categorías ordenadas, seleccione los correspondientes a la **Tau-b de Kendall** o **Spearman**, que miden la asociación entre órdenes de rangos. Los coeficientes de correlación pueden estar entre -1 (una relación negativa perfecta) y +1 (una relación positiva perfecta). Un valor 0 indica que no existe una relación lineal. Al interpretar los resultados, se debe evitar extraer conclusiones de causa-efecto a partir de una correlación significativa.

   Prueba de significación
   Puede seleccionar probabilidades bilaterales o unilaterales. Si conoce de antemano la dirección de la asociación, seleccione **Unilateral**. De lo contrario, seleccione **Bilateral**.

   Señalar las correlaciones significativas
   Los coeficientes de correlación significativos al nivel 0,05 se identifican por medio de un solo asterisco y los significativos al nivel 0,01 se identifican con dos asteriscos.

   Mostrar sólo el triángulo inferior
   Cuando se selecciona, sólo se presenta en el resultado el triángulo inferior de la tabla de matriz de correlaciones. Cuando no se selecciona, se presenta la tabla de matriz de correlaciones completa en la salida. El valor permite que la salida de la tabla cumpla las directrices de estilo APA.

   Mostrar diagonal
   Cuando se selecciona, el triángulo inferior de la tabla de matriz de correlaciones con los valores diagonales se presenta en el resultado. El valor permite que la salida de la tabla cumpla las directrices de estilo APA.

2. De manera opcional, puede seleccionar lo siguiente:

   - Haz clic en **Opciones...** para especificar las estadísticas de correlación de Pearson y la configuración de los valores perdidos.
   - Haz clic en **Estilo...** para especificar condiciones para cambiar automáticamente las propiedades de las tablas pivotantes basándose en condiciones específicas.
   - Haga clic en **Bootstrap...** para derivar estimaciones robustas de errores estándar e intervalos de confianza para estimaciones como la media, la mediana, la proporción, la odds ratio, el coeficiente de correlación o el coeficiente de regresión.
   - Haga clic en **Intervalo de confianza...** para configurar las opciones de estimación de los intervalos de confianza.

Correlaciones bivariadas: Opciones

Estadísticos
Para correlaciones de Pearson, puede elegir una o ambas de estas opciones:

Medias y desviaciones estándar
Se muestran para cada variable. También se muestra el número de casos con valores no perdidos. Los valores perdidos se consideran según cada variable individual, sin tener en cuenta la opción elegida para la manipulación de los valores perdidos.

Desviaciones de productos vectoriales y covarianzas
Se muestran para cada pareja de variables. El producto vectorial de las desviaciones es igual a la suma de los productos de las variables corregidas respecto a la media. Éste es el numerador del coeficiente de correlación de Pearson. La covarianza es una medida no estandarizada de la relación entre dos variables, igual a la desviación del producto vectorial dividido por *N*-1.

Valores perdidos
Puede elegir uno de los siguientes:

Excluir casos según pareja
Se excluyen del análisis los casos con valores perdidos para una o ambas variables de la pareja que forma un coeficiente de correlación. Debido a que cada coeficiente está basado en todos los casos que tienen códigos válidos para esa pareja concreta de variables, en cada cálculo se utiliza la mayor cantidad de información disponible. Esto puede dar como resultado un grupo de coeficientes basados en un número de casos variable.

Excluir casos según lista
Se excluyen de todas las correlaciones los casos con valores perdidos para cualquier variable.

Intervalo de confianza de correlaciones bivariadas

El diálogo Intervalo de confianza proporciona opciones para la estimación de los intervalos de confianza. El diálogo está disponible cuando se selecciona **Pearson**, **Tau-b de Kendall** o **Spearman** en el diálogo Correlaciones bivariadas.

Estimar el intervalo de confianza del parámetro de correlación bivariada
Controla la estimación de intervalo de confianza del parámetro de correlación bivariada. Cuando se selecciona, se produce la estimación del intervalo de confianza.

Intervalo de confianza (%)
Especifica el nivel de confianza para todos los intervalos de confianza generados. Especifique un valor numérico entre 0 y 100. 95 es el valor predeterminado.

Correlación de Pearson
El valor **Aplicar el ajuste de sesgo** controla si se aplica el ajuste de sesgo. De forma predeterminada, el valor no está seleccionado, por lo que no se tiene en cuenta el término de sesgo. Cuando se selecciona, se aplica el ajuste de sesgo a la estimación de los límites de confianza. El valor está disponible cuando se selecciona **Pearson** en el diálogo Correlaciones bivariadas.

Correlación de Spearman
El valor está disponible cuando se selecciona **Spearman** en el diálogo Correlaciones bivariadas y proporciona opciones para estimar la varianza de la correlación de Spearman a través de los métodos siguientes:

- **Fieller, Hartley y Pearson**
- **Bonett y Wright**
- **Coruso y Cliff**

Características adicionales de los comandos CORRELATIONS y NONPAR CORR

La sintaxis de comandos también le permite:

- Escribir una matriz de correlaciones para correlaciones de Pearson que pueda ser utilizada en lugar de los datos en bruto, con el fin de obtener otros análisis como el análisis factorial (con el subcomando `MATRIX`).
- Obtener correlaciones de todas las variables de una lista con todas las variables de una segunda lista (utilizando la palabra clave `WITH` en el subcomando `VARIABLES`).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Correlación de proximidad

Esta documentación guía a los usuarios a través de la especificación de los datos de entrada, la selección del modelo, las funciones de transformación, los ajustes de dimensionalidad y la inclusión de atributos y propiedades en los mapas de proximidad.

La cartografía de proximidad es una técnica de visualización que se utiliza para reducir la dimensionalidad de los datos multivariantes y mostrar las relaciones entre objetos (casos, artículos u otras entidades) en una configuración espacial. Desarrollado originalmente con el nombre de escalado multidimensional (MDS) en psicología, el método utiliza las proximidades (similitudes, disimilitudes o distancias) para situar los objetos de modo que su disposición espacial refleje la estructura de los datos. Los objetos similares aparecen cerca unos de otros; los no similares se colocan más separados. La cartografía de proximidad facilita la interpretación y el reconocimiento de patrones transformando datos relacionales complejos en gráficos intuitivos bidimensionales o tridimensionales.

En IBM SPSS Estadística , el mapeo de proximidad se implementa a través del procedimiento PROXMAP. A diferencia de PROXSCAL, que se limita a analizar matrices de proximidad, PROXMAP acepta múltiples formas de entrada y proporciona una mayor flexibilidad. Admite múltiples fuentes de proximidad, incorpora variables adicionales (atributos y propiedades) y ofrece una amplia gama de opciones de transformación y restricción. Estas características hacen de PROXMAP una potente herramienta para explorar la estructura multivariante e integrar diversos tipos de datos en una representación espacial unificada.

Cuando se dispone de información de proximidad procedente de múltiples fuentes, especificadas directamente o derivadas de datos multivariantes, PROXMAP puede ajustar varios modelos de diferencias individuales además del *modelo de identidad* por defecto. Los modelos disponibles incluyen el modelo de identidad ponderada (dilatación), el modelo euclidiano ponderado (diagonal) y el modelo euclidiano generalizado con restricciones de rango opcionales.

En comparación con las técnicas tradicionales de reducción de dimensiones, como el análisis de componentes principales (ACP) y el MDS clásico, PROXMAP ofrece una representación más directa y flexible de las proximidades. El MDS clásico puede utilizarse para inicializar la configuración. Cuando se aplica a datos multivariantes, el MDS clásico equivale al análisis de coordenadas principales, una variante del ACP.

PROXMAP acepta los siguientes tipos de entrada.

Tipo de entrada	**Descripción**
Proximidades brutas y proximidades derivadas	Las proximidades derivadas se calculan a partir de variables numéricas, ordinales o nominales. Ambos tipos de proximidades se transforman opcionalmente mediante funciones monotónicas (por ejemplo, ordinales o basadas en splines), antes de aproximarlas mediante distancias.
Atributos	Variables que limitan la configuración y aparecen como direcciones en el espacio articular (mapeo supervisado).
Propiedades	Variables suplementarias utilizadas para la interpretación; aparecen en el espacio pero no influyen en la configuración (mapeo no supervisado).

Una característica única de PROXMAP es la posibilidad de utilizar el mismo conjunto de variables en diferentes funciones:

- Como base para la derivación por proximidad
- Como atributos que conforman la configuración
- Como propiedades que apoyan la interpretación

Como alternativa, se pueden asignar distintos conjuntos de variables a cada función.

Además de la transformación lineal, PROXMAP proporciona varias opciones de transformación (incluidas las transformaciones ordinal, spline monotónica, spline no monotónica y nominal) tanto para los atributos

como para las propiedades. Para las proximidades, las transformaciones deben ser monótonas e incluir también transformaciones de potencia. Los resultados incluyen diagnósticos gráficos como diagramas de tensiones, árboles de expansión mínima y gráficos de vecindad para facilitar la interpretación y la evaluación.

Ejemplo

El mapeo de proximidad puede aplicarse ampliamente en diversas disciplinas, como la psicología y las ciencias del comportamiento (percepción, memoria, preferencias), el marketing (mapeo de marcas, segmentación), la sociología (mapeo ideológico, redes sociales), la educación y la psicometría (diagnóstico de pruebas, perfilado de habilidades), la lingüística (similitud semántica, mapeo dialectal), Geografía (distancias percibidas, análisis espacial ecológico), Genómica y Bioinformática (patrones de expresión, estructura de poblaciones, exploración de la heterogeneidad de la expresión unicelular), Quimioinformática (similitud molecular), Informática (visualización de incrustaciones), Antropología y Arqueología (similitud de artefactos, datos craneales).

Datos

Los datos que se van a analizar incluyen variables que, o bien representan una matriz (o matrices) de proximidad, o bien representan datos multivariantes que se convierten en matriz (o matrices) de proximidad.

En el caso de los datos multivariantes, las variables pueden ser una mezcla de variables de escala, ordinales y de nivel de medición nominal. El número mínimo de variables necesarias es 3

Para los datos de proximidad, proporcione el mismo número de variables que de casos.

Tanto para los datos de proximidad como para los datos multivariantes, el número mínimo de casos necesarios es 3.

El resultado incluye configuraciones espaciales, varios gráficos de diagnóstico y estadísticas de ajuste. Los biplots muestran objetos junto a variables, atributos y propiedades en un espacio común.

Para iniciar el análisis PROXMAP, haga clic en **Analice** > **Cartografía** > **Mapas de proximidad**.

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

Cartografía de proximidad: Datos

Con la pestaña **Datos**, puede definir una o varias fuentes de proximidad basadas en variables del conjunto de datos. Cada fuente se crea seleccionando variables del conjunto de datos activo y especificando si representan matrices de proximidad o datos multivariantes de los que se derivarán las proximidades.

Variables

La lista **Variables** muestra todas las variables del conjunto de datos activo. Las variables incluyen variables numéricas, ordinales y nominales que pueden utilizarse en el cálculo de la proximidad. Selecciona una o más variables y muévelas a **Variables en Origen**.

Variables en origen

Variables en la fuente enumera las variables seleccionadas actualmente para la fuente de proximidad activa. Estas variables definen la estructura de la fuente y se utilizan para derivar proximidades (para datos multivariantes) o para especificar una matriz de proximidad (si las proximidades se proporcionan directamente). Utilice los botones de flecha arriba o abajo para reordenar las variables o eliminarlas según sea necesario.

Etiquetas de objeto

Con **Etiquetas de objetos**, puede especificar una variable que proporcione etiquetas para los objetos (casos) que se mostrarán en los gráficos y tablas de salida. Seleccione una variable de la lista y muévala a **Etiquetas de objeto** para asignarla.

definición de origen

A la derecha de las **Variables en Origen**, los campos Nombre de origen y Etiqueta de origen definen el identificador interno y la etiqueta descriptiva de la proximidad actual source.The **El nombre de origen** debe ser un identificador único precedido de *$*.

Utilice los siguientes botones para gestionar las fuentes.

- **Añadir fuente** crea una nueva fuente de proximidad a partir de las variables del panel Variables en fuente.
- **Cambiar fuente** modifica la fuente actualmente seleccionada utilizando la selección de variables actualizada.
- **Eliminar fuente** elimina de la lista la fuente actualmente seleccionada.

Tipo de datos

Utilice el selector **Tipo de datos** para especificar cómo se interpretan las variables seleccionadas. Con el grupo **Tipo de datos**, puede indicar cómo se derivan las proximidades de las variables seleccionadas:

- **Datos de** proximidad Utilice esta opción si su entrada ya es una matriz de proximidad.
- **Datos multivariantes** (por defecto) trata las variables seleccionadas como características de los objetos. PROXMAP obtiene las proximidades calculando las distancias entre los objetos en función de estas variables.
  - **Euclídea** : Las proximidades se basan en distancias euclidianas no normalizadas.
  - **Euclídea normalizada** : Las variables se estandarizan antes de calcular las distancias (por defecto).

Ponderaciones

Cuando **el tipo de datos** se establece en **Datos de proximidad**, se dispone de una lista opcional de **pesos de proximidad**. Aquí puede especificar variables de peso adicionales que se asociarán a los valores de proximidad. Estos pesos ajustan la influencia de cada proximidad en el proceso de optimización.

Cartografía de proximidad: Pesos

Con la pestaña **Pesos**, puede especificar los pesos de proximidad (celda) para cada fuente y los pesos de objeto (caso) para los datos de proximidad. Estas ponderaciones influyen en la forma en que las proximidades contribuyen al proceso de optimización. Esta pestaña sólo es relevante cuando se selecciona el **tipo de Datos**, **Datos de proximidad** en la pestaña **Datos**.

Variables

La lista **Variables** muestra todas las variables del conjunto de datos activo. Estas variables deben contener los pesos de proximidad. Selecciona una o más variables y muévelas a **Pesos en Origen**. Estas variables deben contener los pesos de proximidad para una fuente específica.

Pesos en origen

Pesos en la fuente muestra las variables de peso seleccionadas que están asociadas a la fuente activa en ese momento. Estas variables definen una matriz de pesos de proximidad que coincide con la estructura de la matriz de proximidad correspondiente. El número de variables de peso debe ser igual al número de variables de la matriz de origen.

Utilice **Añadir Pesos**, **Cambiar Pesos** o **Eliminar Pesos** para gestionar las definiciones de peso de la fuente seleccionada en **Fuentes**.

Orígenes

La lista **Fuente(s)** muestra todas las fuentes de proximidad definidas a las que se pueden asignar pesos. Seleccione una fuente para ver o modificar sus pesos asociados.

Ponderaciones de objeto (caso)

Con el campo Etiquetas de objeto bajo los paneles de selección, puede definir una variable que contenga pesos para casos individuales (objetos). Estos pesos ajustan la influencia de cada objeto en la función de pérdida cuando se utilizan datos de proximidad. Las ponderaciones de casos no son aplicables a los datos multivariantes.

Función de ponderaciones

La **función ponderaciones** ofrece tres opciones para esquemas especiales de ponderación de proximidad que son funciones de los valores de proximidad:

- **Ninguna** : No se aplica ninguna ponderación especial. Sólo se utilizan los pesos de proximidad y de objeto (si los hay) especificados por el usuario.
- **Sammon** : Aplica pesos a la proximidad entre los objetos *i* y *j*. Este peso enfatiza las proximidades más pequeñas, como en el mapeo de Sammon.
- **Energía** : Aplica pesos que ponen aún más énfasis en las proximidades más pequeñas.

El peso total aplicado a cada proximidad es el producto de los pesos de proximidad definidos por el usuario, los pesos de caso y la función de peso especial. Estas ponderaciones derivadas rigen la importancia relativa de cada proximidad en la función de pérdida minimizada por PROXMAP.

Cartografía de proximidad: Inicio

La pestaña **Inicial** especifica la configuración inicial para el proceso de optimización iterativo que se utiliza en PROXMAP. Una buena configuración inicial puede acelerar la convergencia y mejorar la calidad de la solución. Puede elegir entre la inicialización automática utilizando el escalado clásico en las proximidades brutas o a sus órdenes de rango. También puede proporcionar una configuración personalizada a partir del conjunto de datos.

Escalado clásico
Esta opción es la predeterminada. Se calcula una solución de escala clásica (Análisis de Coordenadas Principales) basada en las proximidades. Proporciona una configuración inicial derivada analíticamente en el número de dimensiones especificado.

Dos opciones adicionales controlan cómo se tratan las proximidades durante el escalado clásico:

- Con **Valor constante aditivo**, puede especificar una constante fija que se añadirá a todas las proximidades antes de escalar. Esto puede mejorar las propiedades métricas y evitar los valores propios negativos. El valor no debe dar lugar a proximidades negativas.
- Con la opción **Estimar la constante aditiva**, PROXMAP estima automáticamente la constante óptima a añadir. Esto garantiza que la matriz de distancia resultante sea semidefinida positiva.
- Con **Utilizar números de rango para proximidades** puede aplicar el escalado clásico a los órdenes de rango de las proximidades en lugar de a los valores brutos. Así se obtiene una configuración de inicio `ordinal` y se reduce la sensibilidad a los valores atípicos.

Configuración personalizada
Con la **Configuración Personalizada**, puede proporcionar directamente sus propias coordenadas iniciales. Las variables del conjunto de datos se seleccionan como dimensiones de la configuración inicial. Estas variables deben contener valores numéricos válidos sin datos que falten.

La lista de **variables** contiene todas las variables disponibles. Seleccione las variables que contienen la configuración personalizada y muévalas a **Dimensión(es**).

La lista **Dimensiones** muestra las variables seleccionadas para su uso como coordenadas iniciales. El número de variables seleccionadas debe coincidir con el número de dimensiones. Si se proporcionan más variables, los valores se proyectan en los primeros *p* componentes principales, donde *p* es el número de dimensiones.

Número actual de dimensiones muestra la dimensionalidad especificada para la solución final (especificada en la pestaña **Criterios**).

Cartografía de proximidad: Modelo

Con la pestaña **Modelo**, puede especificar el modelo de diferencias individuales que se aplicará al analizar múltiples fuentes de proximidad. También define cómo se gestionan las transformaciones de proximidad entre las fuentes: de manera uniforme (incondicional) o independiente (matriz-condicional).

Modelo de tres vías
Los ajustes en el **modelo de tres vías** determinan cómo se mapean las proximidades de múltiples fuentes en una representación espacial común, mediante el uso de pesos espaciales. El modelo determina cómo se aplican las ponderaciones espaciales al espacio común. Para algunos modelos,

los pesos espaciales son equivalentes a los pesos espaciales. Cuando sólo se define una única fuente, se desactivan todas las opciones excepto **Identidad**.

Opción	Descripción
Identidad	Asume que todas las fuentes comparten la misma configuración sin escalado específico de fuente. Sólo disponible cuando se especifica una única fuente.
Dilatación (identidad ponderada)	Cada fuente se modela con un único factor de escala aplicado uniformemente en todas las dimensiones. Con esta opción, cada fuente puede percibir el espacio común a una escala global diferente.
Euclídeo ponderado (Diagonal)	Cada fuente asigna pesos diferentes a las dimensiones del espacio común. Esto permite a las fuentes destacar diferentes características espaciales.
Euclídeo generalizado	A cada fuente se le permite una versión rotada y reponderada (a escala) del espacio común. Es el modelo más flexible y se adapta a una amplia gama de diferencias individuales. • **Rango completo** : Cada fuente utiliza toda la dimensionalidad del espacio común. • **Rango reducido** : Cuando se aplica una restricción de rango reducido, la configuración de cada fuente (espacio individual, conrango = número de dimensiones - 1) se representa en un subespacio de dimensión inferior del espacio común. Este enfoque implica una rotación ortogonal de la configuración común, seguida de una proyección en un subespacio. Aquí, la rotación y la proyección forman conjuntamente los llamados pesos espaciales.

Condicionalidad / Transformación

Este panel controla cómo se transforman las proximidades antes de ajustar las distancias en la configuración.

- Con la opción **Incondicional**, puede aplicar una única función de transformación a todas las fuentes. Esta opción es adecuada cuando se supone que las fuentes reflejan una estructura común.

  La transformación por defecto se muestra en **Transformación** (ordinal con empates permitidos a desempatar).

  Haga clic en **Definir transformación** para abrir un cuadro de diálogo en el que podrá elegir entre todos los tipos de transformación (Lineal, Potencia, Monotónica, Spline, Ordinal), el tratamiento de los empates y otros parámetros.
- Con la opción **Matriz-Condicional** puede elegir diferentes funciones de transformación para diferentes fuentes o grupos de fuentes. Cuando se selecciona, cada fuente aparece con la función de transformación por defecto.

  Haga clic en **Definir transformación** para elegir otra función que se aplicará a las fuentes seleccionadas. Esta opción sólo se activa cuando se define más de una fuente.

En **Definir transformación**, puede seleccionar cómo se transforman óptimamente las proximidades antes de hacerlas coincidir con las distancias en la configuración espacial. Se encuentran disponibles las siguientes opciones:

Opción	Descripción
Ninguna	No se aplica ninguna transformación. Las proximidades se utilizan tal cual.
Lineal	Aplica una transformación rectilínea a las proximidades. **Interceptar** (casilla de verificación) es añadir una constante a la transformación, lo que permite modelar a escala de intervalo. Si no está marcada, la transformación es a escala proporcional (no se añade ninguna constante).
Potencia	Aplica una transformación de potencia. **Interceptar** (casilla de verificación) es para añadir un desplazamiento constante a la función de potencia, si está activada.
Spline	Ajusta una función polinómica suave (spline) a las proximidades. **El intercepto** incluye un término constante en la base del spline. **El grado** establece la complejidad de los segmentos polinómicos (por ejemplo, 1 = lineal, 2 = cuadrático). **Nudos interiores** controla el número de puntos interiores utilizados para definir la forma de la spline.
Secuencia de nudos	Intervalos iguales: Los nudos están espaciados uniformemente a lo largo del intervalo de proximidad. Percentiles: Los nodos se colocan en cuantiles empíricos de la distribución de proximidad.
Ordinal	Aplica una transformación que preserva el rango (monótona). **Permitir que se desaten los vínculos** : Las proximidades vinculadas pueden transformarse en valores diferentes. **Mantener atadas las proximidades** : Las proximidades vinculadas permanecen vinculadas tras la transformación.

Cartografía de proximidad: Restricciones

Con la pestaña **Restricciones**, puede restringir el espacio de soluciones especificando variables de atributo. Estas variables guían activamente la configuración restringiendo el espacio común de los objetos para que sea una combinación lineal de los atributos. De este modo, es posible realizar un mapeo supervisado, lo que mejora la interpretabilidad al vincular las dimensiones a variables externas significativas.

Seleccione una de las siguientes opciones.

Ninguno
No se imponen restricciones. La configuración viene determinada únicamente por las proximidades (derivadas).

Atributos
Permite la asignación supervisada utilizando una o más variables externas como restricciones de atributos. Cuando se selecciona esta opción, puede definir qué variables actúan como atributos y especificar cómo se transforma cada una.

Variables

La lista **Variables** de la izquierda muestra todas las variables disponibles en el conjunto de datos. Selecciona una o más variables y haz clic en el botón de flecha para moverlas a **Atributos**.

Atributos

El panel derecho muestra todas las variables de atributo seleccionadas junto con la función de transformación por defecto. Estas transformaciones controlan cómo se escalan o transforman los atributos para su uso en la configuración restringida.

Haga clic en **Definir transformación** para asignar una función de transformación al atributo seleccionado. Se encuentran disponibles las siguientes opciones:

- **Lineal** mantiene el atributo en su escala normalizada.
- **Spline** aplica una transformación suave, monotónica o no monotónica, utilizando una función polinómica a trozos.
- **El ordinal** sólo conserva el orden.
- **Nominal** trata las categorías como no ordenadas y las ajusta de forma óptima. Proporciona una única transformación (a diferencia de la transformación nominal para propiedades).

Puede seleccionar varias filas en el campo **Atributos** para definir la misma transformación en varios atributos.

El espacio común se estima bajo la restricción de que las coordenadas del objeto se encuentran dentro del subespacio abarcado por los atributos transformados. El campo **Número actual de dimensiones** muestra la dimensionalidad de la solución, que se establece en la pestaña **Criterios**. Seleccione siempre un número de atributos mayor que el número de dimensiones.

Cartografía de proximidad: Suplemento

Con la pestaña **Suplementos**, puede definir variables de propiedades que ayuden a interpretar la configuración espacial pero que no influyan en su estimación (cartografía no supervisada).

Variables
La lista **Variables** de la izquierda muestra todas las variables disponibles en el conjunto de datos activo. Seleccione una o varias variables que desee tratar como propiedades y muévalas a la lista **Propiedades**. Se puede seleccionar cualquier número de propiedades.

Propiedades

El panel **Propiedades** enumera todas las variables designadas como variables de propiedad, junto con la transformación asignada a cada una.

Haga clic en **Definir transformación** para asignar una función de transformación a una o varias variables de propiedad seleccionadas. Consulte las siguientes opciones disponibles:

- **Linear** mantiene las propiedades en su escala normalizada.
- **Spline** aplica una transformación suave, monotónica o no monotónica, utilizando una función polinómica a trozos.
- **El ordinal** sólo conserva el orden.

- Categorías de ajustes **nominales** como los centroides de los puntos de objeto asociados en la configuración.

Cartografía de proximidad: Criterios

La pestaña **Criterios** especifica los ajustes algorítmicos que controlan la dimensionalidad de la solución, el comportamiento de convergencia del proceso de optimización y la normalización y orientación posteriores a la solución de la configuración resultante.

Dimensionalidad

Establece el número de dimensiones en el espacio común. El valor debe ser igual o superior a 2 e inferior anúmero de objetos - 1. Aumentar el número de dimensiones puede mejorar el ajuste, pero eso puede reducir la interpretabilidad.

Convergencia de algoritmos

Estos ajustes determinan cuándo se detiene la optimización iterativa.

Campo	Descripción	Valor predeterminado
Mínimo de función	Establece el umbral del valor de la función de pérdida por debajo del cual se acepta la solución. Si el valor de la función de pérdida cae por debajo de este mínimo, la iteración se detiene. El valor debe ser mayor o igual que cero.	1.0×10^{-6}
Diferencia de función	Establece el criterio de convergencia basado en el cambio de la función de pérdida entre iteraciones. Si la diferencia absoluta de los valores de las funciones cae por debajo de este valor, se considera que el algoritmo ha convergido. El valor debe ser mayor o igual que cero.	1.0×10^{-12}
Diferencia de configuración	Establece el umbral de convergencia para los cambios en la configuración entre iteraciones. Si el movimiento total de las posiciones de los objetos cae por debajo de este valor, el algoritmo se detiene. El valor debe ser mayor o igual que cero.	1.0×10^{-6}
Número máximo de iteraciones	Especifica el número máximo de iteraciones que puede realizar el algoritmo. Si no se alcanza la convergencia antes de este número, la iteración se detiene de todos modos. El valor debe ser mayor o igual que cero.	1000

Estandarización

Especifica cómo se estandarizan las variables multivariantes (si se utilizan para derivar proximidades) antes del cálculo de la distancia. Esta opción también se aplica a las variables de atributo y propiedad.

- Por defecto se selecciona **Suma de cuadrados = N**.

- Seleccione **Suma de cuadrados = N–1** si desea coherencia con otros procedimientos estadísticos.

Dimensión

Especifica cómo se orienta la configuración del espacio común en el espacio de coordenadas.

Botón de selección	Descripción
Ninguno/Tal cual	No se aplican restricciones. La orientación es en ejes principales, donde los signos de las dimensiones son arbitrarios.
Correlación positiva entre dimensiones	Establece el reflejo de la configuración para maximizar la correlación positiva entre las dimensiones y los datos.
Cuadrante positivo para la mayor coordenada absoluta	Refleja las dimensiones para garantizar que el objeto con la coordenada mayor aparezca en el cuadrante positivo del espacio.
Por dimensión	Puede establecer reflexiones individualmente para cada dimensión. Seleccione la casilla situada junto a una dimensión para aplicar la reflexión a ese eje específico.

Cartografía de proximidad: Salida

Con la subpestaña **Salida**, puede solicitar tablas y gráficos que resuman los datos y resultados del análisis PROXMAP. Contiene subpestañas para cada una de las pestañas principales de análisis (**Datos**, **Inicial**, **Modelo**, **Restricciones**, **Suplementario** y **Criterios**)

Puede solicitar resultados relacionados con cualquier aspecto del análisis. Cada subpestaña ofrece casillas de verificación para solicitar tablas o gráficos específicos.

Datos

En la pestaña **Datos**, puede solicitar resultados relacionados con los datos: datos multivariantes, proximidades y ponderaciones.

Tablas (Datos)

Opción	Descripción
Datos multivariantes originales	Muestra los valores brutos de las variables que se utilizan para derivar las proximidades.
Proximidades originales	Muestra las proximidades proporcionadas directamente por el usuario (si procede).
Proximidades derivadas	Muestra las proximidades calculadas a partir de los datos multivariantes para todas las fuentes.
Proximidades combinadas	Muestra las proximidades agregadas si se combinan varias fuentes.

Tablas (Pesos)

Opción	Descripción
Ponderaciones originales	Enumera los pesos de proximidad brutos (proporcionados por el usuario) que se utilizan para cada fuente.

Opción	Descripción
Ponderaciones derivadas	Proporciona listas de ponderaciones que se derivan de las ponderaciones brutas, las ponderaciones del caso objeto, los datos que faltan y la función de ponderación aplicada a las proximidades.
Pesos del objeto (caja)	Muestra los pesos aplicados por objeto (si se ha especificado).

Gráficos (datos)

El histograma de proximidades muestra la distribución de los valores de proximidad entre todas las fuentes.

Gráficos (pesos)

El histograma de pesos muestra la distribución de los pesos de proximidad (si se han proporcionado o derivado los pesos).

Inicial

Con la subpestaña **Inicial**, puede solicitar resultados relacionados con la configuración inicial, la descomposición del ajuste y las medidas de diagnóstico tras la estimación inicial.

Tablas

Opción	Descripción
Coordenadas	Muestra las coordenadas de los objetos en la configuración inicial.
Direcciones variables	Muestra las direcciones vectoriales de las variables originales en el espacio de configuración (para datos multivariantes). Da coordenadas múltiples para variables nominales (si están presentes en datos multivariantes).
Direcciones de propiedad	Muestra las direcciones de las variables de propiedad (habilitado sólo cuando las propiedades están definidas). Da coordenadas múltiples para variables nominales (si están presentes en datos multivariantes).
Descomposición de estrés	Informa de la contribución de cada objeto al valor global de la pérdida (tensión).
Diagnósticos	Incluye estadísticas de ajuste y diagnósticos para evaluar la solución inicial. Entre ellos se incluyen los valores propios y los vectores propios para el arranque de escala clásico.

Gráficos

Opción	Descripción
Configuración	Muestra una representación gráfica del espacio objeto.
Diagrama de dispersión biespacial multivariante	Combina los puntos del objeto y las direcciones de los vectores variables en un gráfico conjunto, disponible para datos multivariantes.

Opción	Descripción
Diagrama de dispersión biespacial de propiedades	Visualiza la relación entre objetos y variables de propiedades (disponible cuando se definen propiedades).
Gráfico de sedimentación	Representa gráficamente los valores propios de la solución de escala clásica y las dimensiones restantes.
Gráfico de estrés	Un gráfico de burbujas superpuesto al espacio de configuración del objeto. Cada objeto se representa como un punto rodeado por un círculo. El tamaño de cada círculo refleja la contribución relativa de ese objeto a la Tensión Normalizada global de la configuración.

Modelo

Con la subpestaña **Modelo**, puede solicitar resultados gráficos y tabulares relacionados con los espacios comunes e individuales, las transformaciones específicas de la fuente y las medidas de ajuste del modelo.

Tablas

Opción	Descripción
Coordenadas	Lista las coordenadas de los objetos en el espacio común final.
Direcciones variables	Enumera las direcciones de las variables originales (para datos multivariantes). Da coordenadas múltiples para variables nominales (si están presentes en datos multivariantes).
Ponderaciones de espacio	Muestra los pesos espaciales estimados para cada fuente en modelos que permiten diferencias individuales.

Gráficos (espacio común)

Opción	Descripción
Configuración (seleccionada por defecto)	Muestra la configuración espacial común de los objetos. **La configuración** es un ajuste predeterminado que se incluye en la salida, a menos que seleccione otras opciones.
Diagrama de dispersión biespacial multivariante	Muestra que tanto los datos de configuración como las direcciones son multivariantes.
Diagrama de dispersión biespacial de rango reducido	Muestra las proyecciones específicas de la fuente en el espacio común bidimensional si el modelo incluye espacios individuales de rango reducido con rango 1. Este gráfico combina el espacio común con los espacios individuales en un solo gráfico.
Árbol de expansión mínimo	Conecta todos los objetos de la configuración con la longitud total de arista más corta posible. El MST ayuda a visualizar la estructura subyacente

Opción	Descripción
	de la configuración, destacando las agrupaciones y transiciones basadas en la proximidad.
vecinos	Un grafo de vecinos conecta cada objeto de la configuración con sus k vecinos más próximos, en función de las distancias en el espacio de configuración. Este gráfico resalta la estructura local y ayuda a detectar grupos o valores atípicos en la configuración. Se puede elegir el valor del número de vecinos más próximos.
umbral	Conecta pares de objetos en el **espacio de configuración** cuyas proximidades originales o transformadas son menores o iguales que un umbral especificado. Se puede elegir el valor del umbral.

Gráficos (espacio individual)

Opción	Descripción
Configuración	Traza configuraciones espaciales individuales para fuentes seleccionadas (se utiliza con modelos de diferencias individuales).
Ponderaciones de espacio	Muestra los pesos de dimensión para el modelo de Dilatación y el Euclídeo Ponderado.

Gráficos (medidas de transformación y ajuste)

Opción	Descripción
Transformación	Muestra las proximidades transformadas frente a las proximidades originales para las fuentes seleccionadas.
Residuos	Representa las diferencias entre las proximidades transformadas y las distancias ajustadas frente a las proximidades transformadas.
Ajustar	Muestra las distancias ajustadas frente a las distancias transformadas por fuente.
Shepard	Crea un diagrama de Shepard que compara las proximidades y distancias ajustadas frente a las proximidades originales para las fuentes seleccionadas.

Restricciones

La subpestaña **Restricciones** proporciona resultados relacionados con el uso de variables de atributo como restricciones de configuración (mapeo supervisado). Puede solicitar tablas y visualizaciones para las versiones original, transformada y ajustada de estos atributos.

Atributos de tabla

Opción	Descripción
Atributos originales	Muestra los valores brutos de las variables de atributo seleccionadas.
Atributos transformados	Enumera los valores de los atributos una vez aplicadas las transformaciones óptimas.
Ponderación de los atributos	Muestra las ponderaciones calculadas para los atributos para la restricción del espacio común.
Direcciones de atributos	Enumera las coordenadas vectoriales asociadas a cada atributo en el espacio de articulación.

Gráficos

Seleccione uno o más atributos de la siguiente tabla para generar los siguientes gráficos:

Opción	Descripción
Diagrama de dispersión biespacial	Muestra la configuración del objeto con vectores para las variables de atributo seleccionadas.
Transformación	Muestra la función de transformación que se aplica a cada atributo seleccionado.
Residuos	Muestra la diferencia entre los valores transformados ajustados y los valores transformados de los atributos seleccionados.

Projection Plot muestra una proyección unidimensional de los objetos en el espacio común sobre cada uno de los atributos por separado.

Suplementario

La subpestaña **Suplementario** proporciona resultados relacionados con las variables de propiedades, que se utilizan para la interpretación de la configuración pero no afectan a la optimización (mapeo no supervisado).

Tablas Propiedades

Opción	Descripción
Propiedades originales	Muestra los valores brutos de las variables de propiedad.
Propiedades transformadas	Muestra los valores después de aplicar las transformaciones óptimas.
Direcciones de propiedad	Enumera los vectores direccionales de las propiedades en el espacio de configuración (cuando proceda).

Propiedades de los gráficos

Seleccione una o varias propiedades de la lista para generar los siguientes gráficos:

Opción	Descripción
Diagrama de dispersión biespacial	Muestra la configuración del objeto con vectores para las variables de propiedad seleccionadas.

Opción	Descripción
Tranformación	Muestra la transformación de cada propiedad seleccionada.
Residuos	Traza la diferencia entre los valores transformados - ajustados frente a los valores transformados para las propiedades seleccionadas.

Projection Plot muestra una proyección unidimensional de los objetos del espacio común sobre cada una de las propiedades por separado.

Criterios

La subpestaña **Criterios** proporciona información de diagnóstico y visualizaciones relacionadas con el proceso de estimación iterativo y la convergencia del modelo.

Tablas

Opción	Descripción
Resumen (seleccionado por defecto)	Informa de los valores finales de tensión, el número de iteraciones, el tiempo transcurrido y cualquier mensaje de error. **El resumen** es una configuración predeterminada que se incluye en la salida, a menos que selecciones otras opciones.
Diagnósticos	Enumera otros estadísticos de ajuste, como la fórmula de tensión 1 de Kruskal, el coeficiente de alienación, la dispersión contabilizada, el coeficiente de congruencia de Tucker y las correlaciones de Pearson y Spearman.
Historial de iteraciones	Muestra la tensión normalizada para cada iteración, la diferencia de tensión entre iteraciones y el cambio de configuración (diferencia de coordenadas) entre iteraciones. Para limitar el tamaño de la tabla, la tabla del historial de iteraciones sólo informa de un subconjunto de iteraciones, siguiendo potencias de 2. Esto capta el patrón general de convergencia sin enumerar todas las iteraciones
Descomposición de estrés	Descompone el valor de tensión global en contribuciones de objetos individuales.

Gráficos

Opción	Descripción
Historial de iteraciones	Muestra el gráfico de un subconjunto de los valores de las iteraciones, siguiendo potencias de 2. Valores de la tensión normalizada frente al número de iteración.
Gráfico de estrés	Un gráfico de burbujas que muestra las contribuciones de los objetos en el espacio común mostrando círculos alrededor de los

Opción	Descripción
	puntos de los objetos que son proporcionales a su contribución a la tensión normalizada.

Cartografía de proximidad: Guardar

Con la pestaña **Guardar**, puede exportar los resultados seleccionados a un nuevo conjunto de datos.

Puede elegir guardar los siguientes elementos:

Coordenadas iniciales
Las coordenadas de la configuración inicial (espacio inicial).

Coordenadas finales
Las coordenadas de la configuración final (espacio común).

Coordenadas de modelo
Las coordenadas de las configuraciones individuales (espacios individuales). Esta opción sólo se activa cuando el número de fuentes es superior a uno.

Proximidades derivadas
Las proximidades derivadas de datos multivariantes, o las proximidades originales si se facilitan directamente.

Proximidades combinadas
Las proximidades que se agregan sobre todas las fuentes (si se definen múltiples fuentes).

Proximidades transformadas
Las proximidades después de aplicar las funciones de transformación seleccionadas.

Atributos transformados
Los atributos transformados. La opción sólo se activa si se especifican atributos en la pestaña **Restricciones**.

Atributos ajustados
Los valores de los atributos ajustados a la configuración. La opción sólo se activa si se seleccionan los atributos.

Propiedades transformadas
Los valores transformados de las propiedades suplementarias. La opción sólo está activada si las propiedades están definidas en la pestaña **Suplementos**.

Propiedades ajustadas
Los valores de las propiedades proyectados sobre la configuración. La opción sólo se activa si se selecciona **Propiedades**.

Elegir conjunto de datos

Los elementos seleccionados se escriben en un nuevo conjunto de datos. Haga clic en el botón de elipsis (...) para abrir el cuadro de diálogo **Guardar datos como**, donde puede especificar el nombre del archivo y la carpeta de destino para el nuevo conjunto de datos.

Correlaciones parciales

El procedimiento Correlaciones parciales calcula los coeficientes de correlación parcial, los cuales describen la relación lineal existente entre dos variables mientras se controlan los efectos de una o más variables adicionales. Las correlaciones son medidas de asociación lineal. Dos variables pueden estar perfectamente relacionadas, pero si la relación no es lineal, el coeficiente de correlación no es un estadístico adecuado para medir su asociación.

Ejemplo
¿Existe alguna relación entre la financiación sanitaria y las tasas de enfermedad? Aunque cabe esperar que dicha relación sea negativa, un estudio describe una correlación *positiva* significativa: si la financiación sanitaria aumenta, las tasas de enfermedad parecen disminuir. Sin embargo, si se

controla la tasa de visitas de visitadores médicos, se elimina prácticamente la correlación positiva observada. La financiación sanitaria y las tasas de enfermedad sólo parecen estar relacionadas positivamente debido a que más personas tienen acceso a la sanidad si la financiación aumenta, lo que tiene como resultado que los médicos y hospitales informen de más enfermedades.

Estadísticos
Para cada variable: número de casos sin valores perdidos, media y desviación estándar. Matrices de correlación de orden cero y parcial, con grados de libertad y niveles de significación.

Consideraciones sobre los datos

Datos
Utilice variables cuantitativas y simétricas.

Supuestos
El procedimiento Correlaciones parciales supone que cada par de variables es normal bivariante.

Obtención de correlaciones parciales

1. En los menús seleccione:

   Analizar > **Correlacionar** > **Parcial...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione dos o más variables numéricas para las que se van a calcular las correlaciones parciales.
3. Elija una o más variables numéricas de control.

También se encuentran disponibles las siguientes opciones:

Prueba de significación
Puede seleccionar probabilidades bilaterales o unilaterales. Si conoce de antemano la dirección de la asociación, seleccione **Unilateral**. De lo contrario, seleccione **Bilateral**.

Mostrar el nivel de significación real
De forma predeterminada, se muestran la probabilidad y los grados de libertad para cada coeficiente de correlación. Si anula la selección de este elemento, los coeficientes significativos al nivel 0,05 se identifican con un asterisco, los coeficientes significativos al nivel 0,01 se identifican con un asterisco doble y se eliminan los grados de libertad. Este ajuste afecta a las matrices de correlación parcial y de orden cero.

Correlaciones parciales: Opciones

Estadísticos. Puede elegir una o ambas de las siguientes opciones:

- **Medias y desviaciones estándar.** Se muestran para cada variable. También se muestra el número de casos con valores no perdidos.
- **Correlaciones de orden cero.** Se muestra una matriz de las correlaciones simples entre todas las variables, incluyendo las variables de control.

Valores perdidos. Puede elegir una de las siguientes alternativas:

- **Excluir casos según lista.** Se excluyen de todos los cálculos los casos que presenten valores perdidos para cualquier variable, incluso si es para las variables de control.
- **Excluir casos según pareja.** Para el cálculo de las correlaciones de orden cero, en las que se basan las correlaciones parciales, no se utilizará un caso si tiene valores perdidos en una o ambas variables de un par. La eliminación según pareja aprovecha el máximo de los datos que sean posibles. Sin embargo, el número de casos puede variar de unos coeficientes a otros. Cuando se activa esta opción, los grados de libertad para un coeficiente parcial determinado se basan en el número menor de casos utilizado en el cálculo de cualquiera de las correlaciones de orden cero necesarias para el cálculo de dicho coeficiente parcial.

Características adicionales del comando PARTIAL CORR

La sintaxis de comandos también le permite:

- Leer una matriz de correlaciones de orden cero o escribir una matriz de correlaciones parciales (mediante el subcomando MATRIX).
- Obtener correlaciones parciales entre dos listas de variables (mediante la palabra clave WITH en el subcomando VARIABLES).
- Obtener análisis múltiples (mediante varios subcomandos VARIABLES).
- Especificar otros valores para solicitar (por ejemplo, las correlaciones parciales tanto de primer como de segundo orden) cuando tiene dos variables de control (mediante el subcomando VARIABLES).
- Suprimir coeficientes redundantes (mediante el subcomando FORMAT).
- Mostrar una matriz de correlaciones simples cuando algunos coeficientes no se pueden calcular (mediante el subcomando STATISTICS).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Distancias

Este procedimiento calcula una variedad de estadísticos que miden las similitudes o diferencias (distancias), entre pares de variables o entre pares de casos. Estas medidas de similaridad o de distancia se pueden utilizar después con otros procedimientos, como análisis factorial, análisis de clústeres o escalamiento multidimensional, para ayudar en el análisis de conjuntos de datos complejos.

Ejemplo. ¿Es posible medir similaridades entre pares de automóviles en función de ciertas características, como tipo de motor, consumo y potencia? Al calcular las similitudes entre los coches, se puede obtener una noción de qué coches son similares entre sí y cuáles son diferentes. Para un análisis más formal, puede considerar la aplicación de un análisis jerárquico de clústeres o escalamiento multidimensional a las similitudes para explorar la estructura subyacente.

Estadísticas. Las medidas de diferencia (distancia) para datos de un intervalo son Distancia euclídea, Distancia euclídea al cuadrado, Chebychev, bloque, Minkowski o personalizada; para datos de recuento, medida de chi-cuadrado o phi-cuadrado; para datos binarios, Distancia euclídea, Distancia euclídea al cuadrado, diferencia de tamaño, diferencia de configuración, varianza, forma o Lance y Williams. Las medidas de similitud para datos de intervalos son correlación de Pearson o coseno; para datos binarios, Russel y Rao, concordancia simple, Jaccard, Dice, Rogers y Tanimoto, Sokal y Sneath 1, Sokal y Sneath 2, Sokal y Sneath 3, Kulczynski 1, Kulczynski 2, Sokal y Sneath 4, Hamann, Lambda, *D* de Anderberg, *Y* de Yule, *Q* de Yule, Ochiai, Sokal y Sneath 5, correlación Phi de 4 puntos o dispersión.

Para obtener matrices de distancias

1. En los menús seleccione:

   Analizar > **Correlaciones** > **Distancias...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione al menos una variable numérica para calcular distancias entre casos o seleccione al menos dos variables numéricas para calcular distancias entre variables.
3. Seleccione una alternativa en el grupo Calcular distancias para calcular proximidades entre casos o entre variables.

Distancias: Medidas de disimilaridad

En el grupo Medida, seleccione la alternativa que corresponda al tipo de datos (intervalo, recuento o binario); a continuación, de la lista desplegable, seleccione una las medidas que corresponda a dicho tipo de datos. Las medidas disponibles, por tipo de dato, son:

- **Datos de intervalo.** Distancia euclídea, Distancia euclídea al cuadrado, Chebychev, Bloque, Minkowski o Personalizada.

- **Datos de recuento.** Medida de chi-cuadrado o Medida de phi-cuadrado.
- **Datos binarios.** Distancia euclídea, Distancia euclídea al cuadrado, Diferencia de tamaño, Diferencia de configuración, Varianza, Forma o Lance y Williams. (Introduzca valores para Presente y Ausente para especificar cuáles son los dos valores representativos; las Distancias ignorarán todos los demás valores.)

El grupo Transformar valores permite estandarizar los valores de los datos para casos o variables *antes* de calcular proximidades. Estas transformaciones no se pueden aplicar a los datos binarios. Los métodos disponibles de estandarización son: Puntuaciones *z*, Rango -1 a 1, Rango 0 a 1, Magnitud máxima de 1, Media de 1 o Desviación estándar 1.

El grupo Transformar medidas permite transformar los valores generados por la medida de distancia. Se aplican después de calcular la medida de distancia. Las opciones disponibles son: Valores absolutos, Cambiar el signo y Cambiar la escala al rango 0–1.

Distancias: Medidas de similaridad

En el grupo Medida, seleccione la alternativa que corresponda al tipo de datos (intervalo o binario); a continuación, de la lista desplegable, seleccione una las medidas que corresponda a dicho tipo de datos. Las medidas disponibles, por tipo de dato, son:

- **Datos de intervalo.** Correlación de Pearson o Coseno.
- **Datos binarios.** Russel y Rao, Concordancia simple, Jaccard, Dice, Rogers y Tanimoto, Sokal y Sneath 1, Sokal y Sneath 2, Sokal y Sneath 3, Kulczynski 1, Kulczynski 2, Sokal y Sneath 4, Hamann, Lambda, *D* de Anderberg, *Y* de Yule, *Q* de Yule, Ochiai, Sokal y Sneath 5, Correlación Phi de 4 puntos o Dispersión. (Introduzca valores para Presente y Ausente para especificar cuáles son los dos valores representativos; las Distancias ignorarán todos los demás valores.)

El grupo Transformar valores permite estandarizar los valores de los datos para casos o variables antes de calcular proximidades. Estas transformaciones no se pueden aplicar a los datos binarios. Los métodos disponibles de estandarización son: Puntuaciones *z*, Rango -1 a 1, Rango 0 a 1, Magnitud máxima de 1, Media de 1 y Desviación estándar 1.

El grupo Transformar medidas permite transformar los valores generados por la medida de distancia. Se aplican después de calcular la medida de distancia. Las opciones disponibles son: Valores absolutos, Cambiar el signo y Cambiar la escala al rango 0–1.

Características adicionales del comando PROXIMITIES

El procedimiento Distancias utiliza la sintaxis de comandos PROXIMITIES. La sintaxis de comandos también le permite:

- Especificar cualquier número entero como la potencia para la medida de distancia de Minkowski.
- Especificar cualquier número entero como la potencia y la raíz para una medida de distancia personalizada.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Correlación de distancias

Esta característica requiere la opción Statistics Base.

El procedimiento Correlaciones a distancia calcula la dependencia entre dos variables aleatorias o datos multivariantes. Evalúa la relación no lineal de las variables, a diferencia de los métodos de correlación tradicionales.

Ejemplo

Una empresa minorista quiere saber si los datos demográficos de sus clientes (edad, ingresos, educación, estado civil) influyen en su comportamiento de compra. Los métodos tradicionales de correlación pueden fallar si las relaciones no son lineales o implican interacciones entre múltiples factores.

El procedimiento de Correlación a Distancia puede aplicarse en el escenario con las siguientes ventajas:

- Captura las dependencias no lineales. A diferencia de la correlación de Pearson, la correlación de distancia detecta relaciones aunque no sean lineales.
- Maneja relaciones multidimensionales. Puede evaluar la relación entre una combinación de factores demográficos y las pautas de compra.
- Proporciona una medida exhaustiva. Una correlación de distancia significativa sugiere que los datos demográficos de los clientes influyen mucho en el comportamiento de compra, lo que orienta las estrategias de marketing específicas.

Estadísticas asociadas

- Correlación de distancia (*dCor(X,Y)): Mide la dependencia entre los datos demográficos del cliente (X*) y el comportamiento de compra (*Y*).
- Valor p basado en permutación: Evalúa la significación estadística de la dependencia. Un valor p pequeño (por ejemplo, p<0.05) indica una relación fuerte.
- Covarianza de distancia (*dCov(X,Y)): Cuantifica la fuerza de la dependencia entre variables.*
- Varianza de distancia (*dVar(X), dVar(Y)*): mide la variabilidad dentro de las variables individuales para normalizar la correlación.

Consideraciones sobre los datos

Para el análisis sólo pueden utilizarse variables continuas. Se requiere un mínimo de 2 variables continuas para calcular la correlación de distancias. Se permite un máximo de 30 variables.

Obtención de la correlación de distancias

1. Desde el menú, haga clic en: **Analizar** > **Correlacionar** > **Distancia Correlación**.

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione dos o más variables numéricas para las que se calcula la correlación de distancias. Puede seleccionar hasta 30 variables continuas del conjunto de datos de origen para su análisis.
3. Seleccione una variable identificadora. Si opta por una **Variable ID**, los resultados muestran una tabla de matriz de distancia con IDs.
4. Haga clic en **Aceptar** para ejecutar el procedimiento con la configuración especificada y generar el resultado.

Correlación de distancia: Criterios

El subdiálogo Criterios proporciona parámetros de control para la normalización y otras opciones clave para personalizar el procedimiento de correlación de distancias.

Normalización

Seleccione la opción para asignar un tipo de normalización al conjunto de datos. La normalización estandariza las escalas de las variables, pero puede distorsionar inadvertidamente las relaciones altamente no lineales de los datos, que son cruciales para evaluar con precisión la correlación de distancias. Por tanto, se incluye para facilitar las relaciones lineales.

- La normalización **Mín-Máx** escala cada característica a un rango fijo, normalmente [0, 1]. Esta transformación es útil cuando su conjunto de datos contiene variables con diferentes unidades o rangos. Esta normalización permite una comparación equitativa entre las variables, garantizando que ninguna de ellas domine los cálculos de distancia debido a su escala. Ayuda a mitigar los efectos de los grandes rangos numéricos.
- La normalización de **la puntuación Z** centra los datos en torno a la media con una desviación estándar de 1. Este enfoque es útil para conjuntos de datos que se distribuyen normalmente. Esta normalización reduce la influencia de los valores atípicos, ya que da a cada variable una escala normalizada. Permite una interpretación más fácil de los resultados, sobre todo cuando se

comparan variables diferentes, ya que todas las variables se tratan por igual en relación con su distribución.

- **El escalado robusto** utiliza la mediana y el rango intercuartílico (IQR) para escalar los datos, lo que lo hace resistente a los valores atípicos. Se minimiza la influencia de los valores atípicos, lo que permite realizar cálculos de correlación de distancias más fiables cuando hay valores extremos. Esta técnica es beneficiosa cuando el conjunto de datos contiene anomalías que podrían sesgar los resultados.
- **La transformación logarítmica** está diseñada para reducir la asimetría de los datos, especialmente en las distribuciones con asimetría positiva. Esto puede ayudar a estabilizar la varianza en el conjunto de datos. La transformación logarítmica comprime los valores grandes y expande los pequeños, lo que ayuda a normalizar la distribución. Esto puede mejorar la precisión del análisis de correlación de distancias, especialmente cuando se trabaja con conjuntos de datos muy sesgados.

Intervalo de confianza Porcentaje
Especifique el porcentaje del intervalo de confianza que se utilizará en el análisis estadístico.

El porcentaje del intervalo de confianza indica el nivel de certeza asociado a las estimaciones estadísticas. Un porcentaje más alto refleja una mayor confianza en la precisión del intervalo, pero da lugar a un intervalo más amplio, mientras que un porcentaje más bajo produce un intervalo más estrecho con menos certeza

Permutaciones de la correlación de distancias

En la correlación de distancias, la prueba de permutación se utiliza para evaluar la significación estadística de la correlación de distancias observada entre dos variables. El objetivo es evaluar si la correlación observada es significativamente diferente de lo que podría ocurrir por azar. Uno de los conjuntos de datos (o ambos) se permuta aleatoriamente, rompiendo cualquier asociación existente entre las variables. A continuación, se vuelve a calcular la correlación de distancias para cada permutación, generando una distribución de valores de correlación de distancias bajo la hipótesis nula de ausencia de asociación.

Seleccione **Establecer semilla personalizada** para activar la prueba de permutación y especificar el número de iteraciones y el nivel de significación. Establece el valor en **Seed**. Opcionalmente, puede establecer una semilla para el generador de números aleatorios utilizado en las pruebas de permutación. La semilla debe ser un número entero positivo o cero. El valor por defecto es 2000000.

Establezca el **máximo de iteraciones**. El valor debe ser un entero positivo. El valor por defecto es 100 y no se permiten valores negativos.

Al comparar la correlación de distancias observada con la distribución de correlaciones de distancias de conjuntos de datos permutados, se calcula un valor p. El valor p representa la proporción de correlaciones de distancia permutadas que son mayores o iguales que la correlación de distancia observada. Se utiliza un nivel de significación como umbral para determinar la significación estadística. Si el valor p es inferior al nivel de significación, la correlación observada se considera estadísticamente significativa. Establezca el valor del **Nivel de significación** entre 0 y 1. El valor predeterminado es 0,05.

Correlación de distancia: Imprimir

Imprimir detalles
Seleccione **Imprimir detalles** para visualizar una tabla con las siguientes medidas esenciales: el número de observaciones (N), la media, la desviación típica (Std. Desviación), y los valores mínimo y máximo de cada variable. Estas estadísticas proporcionan una visión global del conjunto de datos y permiten comprender mejor las distribuciones y características de las variables que se comparan.

Coeficientes de correlación de distancias
Seleccione **Coeficientes** de correlación de distancias para generar una tabla que muestre los coeficientes de correlación de distancias calculados para el análisis. `dCor Value` muestra el coeficiente de correlación a distancia entre las variables. Si opta por **Establecer semilla personalizada** en el cuadro de diálogo secundario **Criterios** y establece un nivel de significación,

puede ver la columna `Sig.` que indica el nivel de significación (valor p) asociado al coeficiente de correlación calculado.

`Number of Replicates (R)` indica el número de réplicas (o permutaciones, si se opta por ellas) utilizadas en el análisis.

Estimaciones de la covarianza de distancia
Si los usuarios seleccionan un nivel de intervalo de confianza (IC), la tabla incluye los límites superior e inferior de los intervalos de confianza, lo que proporciona una visión más completa de las estimaciones.

Métricas de distancia y estimaciones de significación
Las estimaciones ayudan a evaluar las relaciones y dependencias entre las variables analizadas.

Matriz de distancias
Seleccione **Matriz** de distancias para generar una tabla que muestre la matriz de distancias para datos unidimensionales (Por ejemplo. variable "Edad"). Se representan las distancias entre cada par de observaciones. Los elementos diagonales muestran distancia cero (distancia a sí mismo). Esta matriz proporciona una visión clara de las distancias entre pares para que pueda realizar un análisis exhaustivo de las relaciones entre las variables. Seleccione la opción para configurar una matriz de distancia. En el cuadro de diálogo **Configurar** matriz de distancias, seleccione variables de la lista para configurar la matriz de distancias.

Guardar matriz de distancias como archivo
Seleccione la opción para guardar las matrices de distancia en un archivo. Indique el nombre del archivo en **Directorio de archivos de salida**. La extensión de archivo admitida es `.sav`.

Correlación de distancias: Gráfico

Seleccione **Trazar** para generar gráficos de Correlación de Distancias. Seleccione **Diagrama** de dispersión para generar un diagrama de dispersión que muestre las distancias entre pares o las correlaciones entre las variables seleccionadas.

Seleccione la variable de la lista que desea representar en el eje X. La variable debe ser mutuamente excluyente con otras variables elegidas para el gráfico de dispersión.

Seleccione otra variable de la lista para trazarla en el eje Y. Esta variable debe ser mutuamente excluyente con otras variables que se elijan para el gráfico de dispersión.

Modelos lineales

Los modelos lineales predicen un destino continuo basándose en relaciones lineales entre el destino y uno o más predictores.

Los modelos lineales son relativamente simples y proporcionan una fórmula matemática fácil de interpretar para la puntuación. Las propiedades de estos modelos se comprenden bien y se pueden crear rápidamente en comparación con el resto de tipos de modelos (como redes neuronales o árboles de decisión) en el mismo conjunto de datos.

Ejemplo. Una correduría de seguros con recursos limitados para investigar las reclamaciones de seguros de los asegurados desea crear un modelo para estimar los costes de las reclamaciones. Al desplegar este modelo a los centros de servicio, los representantes pueden introducir información de la reclamación mientras están al teléfono con un cliente y obtener inmediatamente el coste "esperado" de la reclamación en función de datos de archivo.

Requisitos de campo. Debe haber un objetivo y al menos una entrada. De forma predeterminada, los campos con las funciones predefinidas de Ambos o Ninguno no se utilizan.

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

El objetivo debe ser continuo (escala). No hay ninguna restricción de nivel de medición en los predictores (entradas); los campos categóricos (nominal y ordinal) se utilizan como factores en el modelo y los campos continuos se utilizan como covariables.

Nota: Si un campo categórico tiene más de 1000 categorías, el procedimiento no se ejecuta y no se genera ningún modelo.

Para obtener un modelo lineal

Esta característica requiere la opción Statistics Base.

En los menús seleccione:

Analizar > **Regresión** > **Modelos lineales automáticos...**

1. Asegúrese de que hay al menos un destino y una entrada.
2. Pulse en **Opciones de generación** para especificar cualquier configuración de generación y modelado.
3. Pulse en **Opciones de modelo** para guardar puntuaciones en el conjunto de datos activo y exportar el modelo en un archivo externo.
4. Pulse en **Ejecutar** para ejecutar el procedimiento y crear los objetos Modelo.

Objetivos

¿Cuál es su objetivo principal? Seleccione el objetivo adecuado.

- **Crear un modelo estándar.** El método crea un único modelo para pronosticar el objetivo utilizando los predictores. Por lo general, los modelos estándar son más fáciles de interpretar y pueden ser más rápidos de puntuar que conjuntos de datos potenciados, empaquetados o grandes.
- **Mejorar la precisión del modelo (aumento).** El método crea un modelo de conjunto utilizando potenciación, que genera una secuencia de modelos para obtener predicciones más precisas. Los conjuntos pueden tardar más en construirse y puntuarse que un modelo estándar.

  La potenciación produce una sucesión de "modelos de componente", cada uno creado con el conjunto de datos al completo. Antes de crear cada modelo de componente, los archivos se ponderan basándose en los residuos de los anteriores modelos de componente. Los casos con muchos residuos reciben ponderaciones de análisis relativamente mayores para que el próximo modelo de componente se centre en predecir bien estos archivos. Juntos, estos modelos de componente forman un modelo de conjunto. El modelo de conjunto puntúa nuevos resultados usando una regla de combinación; las reglas disponibles dependen del nivel de medición del objetivo.
- **Mejorar la estabilidad del modelo (bagging).** El método crea un modelo de conjunto utilizando bagging (agregación de simulación de muestreo), que genera modelos múltiples para obtener predicciones más fiables. Los conjuntos pueden tardar más en construirse y puntuarse que un modelo estándar.

  La agregación de simulación de muestreo (bagging) produce réplicas del conjunto de datos de entrenamiento haciendo muestras con sustitución del conjunto de datos original. Así crea muestras de simulación de muestreo del mismo tamaño que el conjunto de datos original. Después se crea un "modelo de componente" en cada réplica. Juntos, estos modelos de componente forman un modelo de conjunto. El modelo de conjunto puntúa nuevos resultados usando una regla de combinación; las reglas disponibles dependen del nivel de medición del objetivo.
- **Cree un modelo para conjuntos de datos muy grandes (se requiere IBM SPSS Estadística Server).** El método crea un modelo de conjunto dividiendo el conjunto de datos en bloques de datos separados. Seleccione esta opción si su conjunto de datos es demasiado grande para generar cualquiera de los modelos anteriores o para la generación incremental de modelos. Esta opción emplea menos tiempo en su generación, pero puede tardar más en puntuarse que un modelo estándar. Esta opción requiere conectividad con IBM SPSS Estadística Server.

Consulte “Conjuntos ” en la página 167 para ver la información de configuración relacionada con boosting, bagging y conjuntos de datos de gran tamañoboosting y bagging.

Conceptos básicos

Preparar automáticamente datos. Esta opción permite el procedimiento de transformar de forma interna el destino y predictores para maximizar el poder predictivo del modelo; las transformaciones

se guardan con el modelo y se aplican a los nuevos datos para su puntuación. Las versiones originales de los campos transformados se excluyen del modelo. De forma predeterminada, se realiza la siguiente preparación automática de datos.

- **Fecha y hora.** Cada predictor de fecha se transforma en un nuevo predictor continuo que contiene el tiempo transcurrido desde una fecha de referencia (01-01-1970). Cada predictor de hora se transforma en un nuevo predictor continuo que contiene el tiempo transcurrido desde una hora de referencia (00:00:00).
- **Ajustar nivel de medición.** Los predictores continuos con menos de 5 valores distintos se reestructuran como predictores ordinales. Los predictores ordinales con más de 10 valores distintos se reestructuran como predictores continuos.
- **Tratamiento de valores atípicos.** Los valores de los predictores continuos que recaen más allá de un valor de corte (3 desviaciones estándar de la media) se establecen con el valor de corte.
- **Gestión de valores perdidos.** Los valores perdidos de los predictores nominales se sustituyen por el modo de la partición de entrenamiento. Los valores perdidos de los predictores ordinales se sustituyen por la mediana de la partición de entrenamiento. Los valores perdidos de los predictores continuos se sustituyen por la media de la partición de entrenamiento.
- **Fusión supervisada.** Hace un modelo más parsimonioso reduciendo el número de campos que deben procesarse junto con el destino. Las categorías similares se identifican en función de la relación entre la entrada y destino. Las categorías que no son significativamente diferentes (es decir, que tienen un valor p superior al valor 0,1) se fusionan. Tenga en cuenta que si todas las categorías se combinan en una, las versiones original y derivada del campo se excluyen del modelo porque no tienen ningún valor como predictor.

Nivel de confianza. Éste es el nivel de confianza que se utiliza para calcular las estimaciones de intervalos de los coeficientes de modelos en la vista Coeficientes. Especifique un valor mayor que 0 y menor que 100. El valor predeterminado es 95%.

Selección de modelos

Método de selección de modelos. Seleccione uno de los métodos de selección de modelos (a continuación se encuentran los detalles) o **Incluya todos los predictores**, que simplemente introduce todos los predictores disponibles como términos del modelo de efectos principales. De forma predeterminada, se utiliza **Pasos sucesivos hacia adelante**.

Selección de Pasos sucesivos hacia adelante. Comienza sin efectos en el modelo y añade y elimina efectos paso por paso hasta que ya no se puedan añadir o eliminar según los criterios de los pasos sucesivos.

- **Criterios para entrada/eliminación.** Éste es el estadístico utilizado para determinar si debe añadirse o eliminarse un efecto del modelo. **Criterio de información (AICC)** se basa en la similitud del conjunto de entrenamiento que se le da al modelo, y se ajusta para penalizar modelos excesivamente complejos. **Estadísticos de F** se utiliza en una prueba estadística de la mejora en el error de modelo. **R cuadrado corregida** se basa en el ajuste del conjunto de entrenamiento, y se ajusta para penalizar modelos excesivamente complejos. **Criterio de prevención sobreajustado (ASE)** se basa en el ajuste (error cuadrado medio o ASE) del conjunto de prevención sobreajustado. El conjunto de prevención sobreajustado es una submuestra aleatoria de aproximadamente el 30% del conjunto de datos original que no se utiliza para enseñar el modelo.

  Si se selecciona otro criterio que no sea **Estadísticos de F**, se añadirá al modelo cada paso del efecto que se corresponda con el aumento positivo mayor en el criterio. Se eliminará cualquier efecto en el modelo que se corresponda con una disminución en el criterio.

  Si se selecciona **Estadísticos de F** como criterio, cada paso en el efecto que tenga el valor *p* más pequeño inferior al umbral especificado, se añadirá **Incluir efectos con valores p inferiores a** al modelo. El valor predeterminado es 0.05. Cualquier efecto en el modelo con un valor *p* superior al umbral especificado, **Eliminar efectos con valores p mayores que**, será eliminado. El valor predeterminado es 0.10.

- **Personalizar número máximo de efectos en el modelo final.** De forma predeterminada, pueden introducirse todos los efectos disponibles en el modelo. Del mismo modo, si el algoritmo por pasos sucesivos termina con un paso con el número máximo de efectos especificado, el algoritmo se detiene con el conjunto actual de efectos.
- **Personalizar número máximo de pasos.** El algoritmo por pasos sucesivos termina tras un cierto número de pasos. De forma predeterminada, es 3 veces el número de efectos disponibles. Del mismo modo, especifique un entero positivo para el número máximo de pasos.

Selección de mejores subconjuntos. Comprueba "todos los modelos posibles", o al menos el subconjunto más grande de los modelos posibles que los pasos sucesivos hacia adelante, para seleccionar el mejor según el criterio de mejores subconjuntos. El **criterio de información (AICC)** se basa en la probabilidad del conjunto de entrenamiento dado el modelo, y se ajusta para penalizar modelos excesivamente complejos. **R cuadrado corregida** se basa en el ajuste del conjunto de entrenamiento, y se ajusta para penalizar modelos excesivamente complejos. **Criterio de prevención sobreajustado (ASE)** se basa en el ajuste (error cuadrado medio o ASE) del conjunto de prevención sobreajustado. El conjunto de prevención sobreajustado es una submuestra aleatoria de aproximadamente el 30% del conjunto de datos original que no se utiliza para enseñar el modelo.

Se selecciona el modelo con el valor mayor del criterio como el mejor modelo.

Nota: La selección de mejores subconjuntos requiere más trabajo computacional que la selección de avanzar por pasos. Cuando los mejores subconjuntos se procesan junto con boosting, bagging y conjuntos de datos de gran tamaño, la generación de un modelo estándar generado mediante una selección de avanzar por pasos puede tardar considerablemente más tiempo.

Conjuntos

Estos ajustes determinan el comportamiento de la agrupación que se produce cuando los conjuntos de datos de gran tamaño o de boosting o bagging son obligatorios en Objetivos. Las opciones no aplicables al objetivo seleccionado se ignorarán.

Bagging y conjuntos de datos muy grandes. Al puntuar un conjunto, ésta es la regla utilizada para combinar los valores pronosticados a partir de los modelos básicos para calcular el valor de puntuación del conjunto.

- **Regla de combinación predeterminada para objetivos continuos.** Los valores pronosticados de conjunto para objetivos continuos pueden combinarse mediante la media o mediana de los valores pronosticados a partir de los modelos básicos.

Tenga en cuenta que cuando el objetivo es mejorar la precisión del modelo, se ignoran las selecciones de reglas de combinación. El boosting siempre utiliza un voto de mayoría ponderada para puntuar objetivos categóricos y una mediana ponderada para puntuar objetivos continuos.

Boosting y bagging. Especifique el número de modelos básicos que debe generarse cuando el objetivo es mejorar la precisión o estabilidad del modelo; en el caso del bagging, se trata del número de muestras de simulación de muestreo. Debe ser un número entero positivo.

Avanzado

Replicar resultados. Al establecer una semilla aleatoria podrá replicar análisis. El generador de números aleatorios se utiliza para seleccionar qué registros están en el conjunto de prevención sobreajustado. Especifique un entero o pulse en **Generar**, lo que creará un entero pseudo-aleatorio entre 1 y 2147483647, ambos inclusive. El valor predeterminado es 54752075.

Opciones de modelos

Guardar valores predichos en el conjunto de datos. El nombre de variable predeterminado es *PredictedValue*.

Exportar modelo. Graba el modelo en un archivo `.zip` externo. Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo. Especifique un nombre

de archivo exclusivo válido. Si la especificación de archivo hace referencia a un archivo existente, se sobrescribirá el archivo.

Resumen del modelo

La vista Resumen de modelos es una instantánea, un resumen de un vistazo del modelo y su ajuste.

Tabla. La tabla identifica algunos ajustes de modelo de alto nivel, incluyendo:

- Nombre del destino especificado en la pestaña Campos,
- Si la preparación automática de los datos se realizó tal como se especificaba en los ajustes de Procedimientos básicos,
- El criterio de selección y el método de selección de modelo especificado en los ajustes de Selección de modelo. También se muestra el valor del criterio de selección del modelo final, y viene presentado en un formato cuanto más pequeño mejor.

Gráfico. El gráfico muestra la precisión del modelo final, que se presenta en el formato mayor es mejor. El valor es $100 \times R^2$ ajustado para el modelo final.

Preparación automática de datos

Esta vista muestra información a cerca de qué campos se excluyen y cómo los campos transformados se derivaron en el paso de preparación automática de datos (ADP). Para cada campo que fue transformado o excluido, la tabla enumera el nombre del campo, su rol en el análisis y la acción tomada por el paso ADP. Los campos se clasifican por orden alfabético ascendente de nombres de campo. Las acciones que puede realizarse para cada campo incluyen:

- **Derivar duración: meses** calcula el tiempo transcurrido en meses a partir de los valores de un campo que contiene las fechas hasta la fecha actual del sistema.
- **Derivar duración: horas** calcula el tiempo transcurrido en horas a partir de los valores de un campo que contiene las horas hasta la hora actual del sistema.
- **Cambiar el nivel de medición de continuo a ordinal** reestructura los campos continuos que tienen menos de 5 valores exclusivos como campos ordinales.
- **Cambiar el nivel de medición de ordinal a continuo** reestructura los campos ordinales que tienen menos de 10 valores exclusivos como campos continuos.
- **Recortar valores atípicos** define los valores de los predictores continuos que recaen más allá de un valor de corte (3 desviaciones estándar de la media) con el valor de corte.
- **Reemplazar los valores perdidos** sustituye los valores perdidos en los campos nominales con el modo, en los campos ordinales con la mediana y en los campos continuos con la media.
- **Fusionar categorías para maximizar la asociación con el destino** identifica las categorías de predictor "similares" en función de la relación entre la entrada y el destino. Las categorías que no son significativamente diferentes (es decir, que tienen un valor *p* superior al valor 0,05) se fusionan.
- **Excluir predictor constante / tras tratamiento de valores atípicos / tras fusionar categorías** elimina los predictores que tienen un único valor, posiblemente después de que se hayan realizado otras acciones de ADP.

Importancia de predictor

Normalmente, desea centrar sus esfuerzos de modelado en los campos del predictor que importan más y considera eliminar o ignorar las que importan menos. El predictor de importancia de la variable le ayuda a hacerlo indicando la importancia relativa de cada predictor en la estimación del modelo. Como los valores son relativos, la suma de los valores de todos los predictor de la visualización es 1.0. La importancia del predictor no está relacionada con la precisión del modelo. Sólo está relacionada con la importancia de cada predictor para realizar una predicción, independientemente de si ésta es precisa o no.

Predicho por observado

Muestra un diagrama de dispersión agrupado de los valores predichos en el eje vertical por los valores observados en el eje horizontal. En teoría, los puntos deberían encontrarse en una línea de 45 grados; esta vista puede indicar si hay registros para los que el modelo realiza una predicción particularmente malo.

Residuos

Muestra un gráfico de diagnosis de los residuos del modelo.

Estilos de gráfico. Existen varios estilos de visualización diferentes, que son accesibles desde la lista desplegable **Estilo**.

- **Histograma.** Se trata de un histograma en intervalos de los residuos estudentizados de una superposición de la distribución normal. Los modelos lineales asumen que los residuos tienen una distribución normal, de forma que el histograma debería estar muy cercano a la línea continua.
- **Gráfico p-p.** Se trata de un gráfico probabilidad-probabilidad en intervalos que compara los residuos estudentizados con una distribución normal. Si la curva de los puntos representados es menos pronunciada que la línea normal, los residuos muestran una variabilidad mayor que una distribución normal; si la curva es más pronunciada, los residuos muestran una variabilidad inferior que una distribución normal. Si los puntos representados tienen una curva con forma en S, la distribución de los residuos es asimétrica.

Valores atípicos

Esta tabla enumera los registros que ejercen una influencia excesiva sobre el modelo, y muestra el ID de registro (si se especifica en la pestaña Campos), el valor objetivo y la distancia de Cook. La distancia de Cook es una medida de cuánto cambiarían los residuos de todos los registros si un registro en particular se excluyera del cálculo de los coeficientes del modelo. Una distancia de Cook grande indica que la exclusión de un registro cambia sustancialmente los coeficientes, y por lo tanto debe considerarse relevante.

Los registros relevantes deben examinarse cuidadosamente para determinar si puede darles menos importancia en la estimación del modelo, truncar los valores atípicos a algún umbral aceptable o eliminar los registros relevantes completamente.

Efectos

Esta vista muestra el tamaño de cada efecto en el modelo.

Estilos. Existen varios estilos de visualización diferentes, que son accesibles desde la lista desplegable **Estilo**.

- **Diagrama.** Es un gráfico en el que los efectos se clasifican desde arriba hacia abajo con una importancia de predictores descendente. Las líneas de conexión del diagrama se ponderan tomando como base la significación del efecto, con un grosor de línea mayor correspondiente a efectos con mayor significación (valores *p* inferiores). Al pasar el ratón sobre una línea de conexión se muestra una ayuda contextual que muestra el valor *p* y la importancia del efecto. Este es el valor predeterminado.
- **Tabla.** Se trata de una tabla ANOVA para el modelo completo y los efectos de modelo individuales. Los efectos individuales se clasifican desde arriba hacia abajo con una importancia de predictores descendente. Tenga en cuenta que, de forma predeterminada, la tabla se contrae para mostrar únicamente los resultados del modelo general. Para ver los resultados de los efectos de modelo individuales, pulse en la casilla **Modelo corregido** de la tabla.

Importancia del predictor. Existe un control deslizante Importancia del predictor que controla qué predictores se muestran en la vista. Esto no cambia el modelo, simplemente le permite centrarse en los predictores más importantes. De forma predeterminada, se muestran los 10 efectos más importantes.

Significación. Existe un control deslizante Significación que controla aún más qué efectos se muestran en la vista, a parte de los que se muestran tomando como base la importancia de predictor. Se ocultan los

efectos con valores de significación superiores al valor del control deslizante. Esto no cambia el modelo, simplemente le permite centrarse en los efectos más importantes. El valor predeterminado es 1.00, de modo que no se filtran efectos tomando como base la significación.

Coeficientes

Esta vista muestra el valor de cada coeficiente en el modelo. Tenga en cuenta que los factores (predictores categóricos) tienen codificación de indicador dentro del modelo, de modo que los **efectos** que contienen los factores generalmente tendrán múltiples **coeficientes** asociados: uno por cada categoría exceptuando la categoría que corresponde al parámetro (referencia) redundante.

Estilos. Existen varios estilos de visualización diferentes, que son accesibles desde la lista desplegable **Estilo**.

- **Diagrama.** Es un gráfico que muestra la intersección primero, y luego clasifica los efectos desde arriba hacia abajo con una importancia de predictores descendente. Dentro de los efectos que contienen factores, los coeficientes se clasifican en orden ascendente de valores de datos. Las líneas de conexión del diagrama se colorean tomando como base el signo del coeficiente (consulte la descripción del diagrama) y se ponderan en función de la significación del coeficiente, con un grosor de línea mayor correspondiente a coeficientes con mayor significación (valores *p* inferiores). Al pasar el ratón sobre una línea de conexión se muestra una ayuda contextual que muestra el valor del coeficiente, su valor *p* y la importancia del efecto con el que está asociado el parámetro. Este es el estilo predeterminado.
- **Tabla.** Muestra los valores, las pruebas de significación y los intervalos de confianza para los coeficientes de modelos individuales. Tras la intersección, los efectos se clasifican desde arriba hacia abajo con una importancia de predictores descendente. Dentro de los efectos que contienen factores, los coeficientes se clasifican en orden ascendente de valores de datos. Tenga en cuenta que, de forma predeterminada, la tabla se contrae para mostrar únicamente los resultados de coeficiente, significación e importancia de cada parámetro de modelo. Para ver el error estándar, el estadístico *t* y el intervalo de confianza, pulse en la casilla **Coeficiente** de la tabla. Si pasa el ratón sobre el nombre de un parámetro de modelo en la tabla aparece una ayuda contextual que muestra el nombre del parámetro, el efecto con el que está asociado y (para los predictores categóricos) las etiquetas de valor asociadas con el parámetro de modelo. Esto puede ser especialmente útil para ver las nuevas categorías creadas cuando la preparación de datos automática fusiona categorías similares de un predictor categórico.

Importancia del predictor. Existe un control deslizante Importancia del predictor que controla qué predictores se muestran en la vista. Esto no cambia el modelo, simplemente le permite centrarse en los predictores más importantes. De forma predeterminada, se muestran los 10 efectos más importantes.

Significación. Existe un control deslizante Significación que controla aún más qué coeficientes se muestran en la vista, a parte de los que se muestran tomando como base la importancia de predictor. Se ocultan los coeficientes con valores de significación superiores al valor del control deslizante. Esto no cambia el modelo, simplemente le permite centrarse en los coeficientes más importantes. El valor predeterminado es 1.00, de modo que no se filtran coeficientes tomando como base la significación.

Medias estimadas

Son gráficos representados para predictores significativos. El gráfico muestra el valor estimado de modelo del objetivo en el eje vertical de cada valor del predictor en el eje horizontal, que alberga el resto de los predictores constantes. Proporciona una visualización útil de los efectos de los coeficientes de cada predictor en el objetivo.

Nota: si no hay predictores significativos, no se generan medias estimadas.

Resumen de generación de modelos

Cuando se selecciona un algoritmo de selección de modelos que no sea **Ninguno**, proporciona algunos detalles del proceso de generación de modelos.

Avanzar por pasos. Cuando la selección por pasos hacia delante es el algoritmo de selección, la tabla muestra los últimos 10 pasos en el algoritmo de selección por pasos hacia delante. Para cada paso, se

muestran el valor del criterio de selección y los efectos en el modelo en ese paso. Esto ofrece el sentido del grado de contribución de cada paso al modelo. Cada columna le permite clasificar las filas, de modo que es posible ver con mayor facilidad qué efectos hay en un paso en particular.

Mejores subconjuntos. Cuando Mejores subconjuntos es el algoritmo de selección, la tabla muestra los 10 modelos principales. Para cada modelo, se muestran el valor del criterio de selección y los efectos en el modelo. Esto ofrece un sentido de la estabilidad de los modelos principales; si tienden a tener muchos efectos similares con pocas diferencias, puede tenerse una confianza casi completa en el modelo "principal"; si tienden a tener muchos efectos diferentes, algunos efectos pueden ser demasiado parecidos y deberían combinarse (o eliminar uno). Cada columna le permite clasificar las filas, de modo que es posible ver con mayor facilidad qué efectos hay en un paso en particular.

Regresión lineal

La regresión lineal estima los coeficientes de la ecuación lineal, con una o más variables independientes, que mejor prediga el valor de la variable dependiente. Por ejemplo, puede intentar predecir el total de ventas anuales de un vendedor (la variable dependiente) a partir de variables independientes tales como la edad, la formación y los años de experiencia.

Ejemplo. ¿Están relacionados el número de partidos ganados por un equipo de baloncesto en una temporada con la media de puntos que el equipo marca por partido? Un diagrama de dispersión indica que estas variables están relacionadas linealmente. El número de partidos ganados y la media de puntos marcados por el equipo adversario también están relacionados linealmente. Estas variables tienen una relación negativa. A medida que el número de partidos ganados aumenta, la media de puntos marcados por el equipo adversario disminuye. Con la regresión lineal es posible modelar la relación entre estas variables. Puede utilizarse un buen modelo para predecir cuántos partidos ganarán los equipos.

Estadísticas. Para cada variable: número de casos válidos, media y desviación estándar. Para cada modelo: coeficientes de regresión, matriz de correlaciones, correlaciones parciales y semiparciales, *R* múltiple, *R* cuadrado, *R* cuadrado corregida, cambio en *R* cuadrado, error estándar de la estimación, tabla de análisis de varianza, valores pronosticados y residuos. Además, intervalos de confianza al 95% para cada coeficiente de regresión, matriz de varianzas-covarianzas, factor de inflación de la varianza, tolerancia, prueba de Durbin-Watson, medidas de distancia (Mahalanobis, Cook y valores de influencia), DfBeta, DfFit, intervalos de predicción e información de diagnóstico por caso. Gráficos: diagramas de dispersión, gráficos parciales, histogramas y gráficos de probabilidad normal.

Regresión lineal: Consideraciones sobre los datos

Datos. Las variables dependiente e independientes deben ser cuantitativas. Las variables categóricas, como la religión, estudios principales o el lugar de residencia, han de recodificarse como variables binarias (dummy) o como otros tipos de variables de contraste.

Supuestos. Para cada valor de la variable independiente, la distribución de la variable dependiente debe ser normal. La varianza de distribución de la variable dependiente debe ser constante para todos los valores de la variable independiente. La relación entre la variable dependiente y cada variable independiente debe ser lineal y todas las observaciones deben ser independientes.

Para obtener un análisis de regresión lineal

1. En los menús seleccione:

   Analizar > **Regresión** > **Lineal...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. En el cuadro de diálogo Regresión lineal, seleccione una variable numérica dependiente.
3. Seleccione una más variables numéricas independientes.

Si lo desea, puede:

- Agrupar variables independientes en bloques y especificar distintos métodos de entrada para diferentes subconjuntos de variables.

- Elegir una variable de selección para limitar el análisis a un subconjunto de casos que tengan valores particulares para esta variable.
- Seleccionar una variable de identificación de casos para identificar los puntos en los diagramas.
- Seleccione una variable numérica de Ponderación MCP para el análisis de mínimos cuadrados ponderados.

MCP. Permite obtener un modelo de mínimos cuadrados ponderados. Los puntos de los datos se ponderan por los inversos de sus varianzas. Esto significa que las observaciones con varianzas grandes tienen menor impacto en el análisis que las observaciones asociadas a varianzas pequeñas. Si el valor de la variable de ponderación es cero, negativo o perdido, el caso queda excluido del análisis.

Métodos de selección de variables en el análisis de regresión lineal

La selección del método permite especificar cómo se introducen las variables independientes en el análisis. Utilizando distintos métodos se pueden construir diversos modelos de regresión a partir del mismo conjunto de variables.

- *Introducir (Regresión)*. Procedimiento para la selección de variables en el que todas las variables de un bloque se introducen en un solo paso.
- *Por pasos*. En cada paso, se especifica la variable independiente que no está en la ecuación que tiene la menor probabilidad de F, si dicha probabilidad es suficientemente pequeña. Las variables ya introducidas en la ecuación de regresión se eliminan de ella si su probabilidad para F llega a ser suficientemente grande. El método termina cuando ya no hay más variables candidatas a ser incluidas o eliminadas.
- *Eliminar*. Procedimiento para la selección de variables en el que las variables de un bloque se eliminan en un solo paso.
- *Eliminación hacia atrás*. Procedimiento de selección de variables en el que todas las variables se introducen en la ecuación y, a continuación, se eliminan secuencialmente. Aquella variable que tenga la menor correlación parcial con la variable dependiente será la primera en ser considerada para su eliminación. Si satisface el criterio de eliminación, se eliminará. Tras haber excluido la primera variable, se pondrá a prueba aquella variable, de las que queden en la ecuación, que presente una correlación parcial más pequeña. El procedimiento termina cuando ya no quedan en la ecuación variables que satisfagan el criterio de eliminación.
- *Selección hacia delante*. Procedimiento de selección de variables por pasos en el que las variables se introducen secuencialmente en el modelo. La primera variable que se considerará introducir en la ecuación será la que tenga mayor correlación, positiva o negativa, con la variable dependiente. Dicha variable se introducirá en la ecuación sólo si cumple el criterio de entrada. Si se introduce la primera variable, a continuación se considerará la variable independiente cuya correlación parcial sea la mayor y que no esté en la ecuación. El procedimiento termina cuando ya no quedan variables que cumplan el criterio de entrada.

Los valores de significación de los resultados se basan en el ajuste de un único modelo. Por ello, estos valores de significación no suelen ser válidos cuando se emplea un método por pasos (pasos sucesivos, hacia delante o hacia atrás).

Todas las variables deben superar el criterio de tolerancia para que puedan ser introducidas en la ecuación, independientemente del método de entrada especificado. El nivel de tolerancia predeterminado es 0,0001. Tampoco se introduce una variable si esto provoca que la tolerancia de otra ya presente en el modelo se sitúe por debajo del criterio de tolerancia.

Todas las variables independientes seleccionadas se añaden a un mismo modelo de regresión. Sin embargo, puede especificar distintos métodos de introducción para diferentes subconjuntos de variables. Por ejemplo, puede introducir en el modelo de regresión un bloque de variables que utilice la selección por pasos sucesivos, y un segundo bloque que emplee la selección hacia adelante. Para añadir un segundo bloque de variables al modelo de regresión, pulse en **Siguiente**.

Regresión lineal: Establecer regla

Los casos definidos por la regla de selección se incluyen en el análisis. Por ejemplo, si selecciona una variable, elija **igual que** y escriba 5 para el valor; de este modo, solamente se incluirán en el análisis los casos para los cuales la variable seleccionada tenga un valor igual a 5. También se permite un valor de cadena.

Regresión lineal: Gráficos

Los gráficos pueden ayudar a validar los supuestos de normalidad, linealidad e igualdad de las varianzas. También son útiles para detectar valores atípicos, observaciones poco usuales y casos de influencia. Tras guardarlos como nuevas variables, dispondrá en el Editor de datos de los valores pronosticados, los residuos y otra información de diagnóstico, con los cuales podrá poder crear gráficos respecto a las variables independientes. Se encuentran disponibles los siguientes gráficos:

Diagramas de dispersión. Puede representar cualquier combinación por parejas de la lista siguiente: la variable dependiente, los valores pronosticados estandarizados, los residuos estandarizados, los residuos eliminados, los valores pronosticados corregidos, los residuos estudentizados o los residuos eliminados estudentizados. Represente los residuos estandarizados frente a los valores pronosticados estandarizados para contrastar la linealidad y la igualdad de las varianzas.

Lista de variables de origen. Lista la variable dependiente (DEPENDNT) y las siguientes variables pronosticadas y de residuos: Valores pronosticados estandarizados (*ZPRED), Residuos estandarizados (*ZRESID), Residuos eliminados (*DRESID), Valores pronosticados corregidos (*ADJPRED), Residuos estudentizados (*SRESID) y Residuos estudentizados eliminados (*SDRESID).

Generar todos los gráficos parciales. Muestra los diagramas de dispersión de los residuos de cada variable independiente y los residuos de la variable dependiente cuando se regresan ambas variables por separado sobre las restantes variables independientes. En la ecuación debe haber al menos dos variables independientes para que se generen los gráficos parciales.

Gráficos de residuos tipificados. Puede obtener histogramas de los residuos tipificados y gráficos de probabilidad normal que comparen la distribución de los residuos tipificados con una distribución normal.

Si se solicita cualquier gráfico, se muestran los estadísticos de resumen para los valores pronosticados estandarizados y los residuos estandarizados (***ZPRED** y ***ZRESID**).

Regresión lineal: almacenamiento de variables nuevas

Puede guardar los valores pronosticados, los residuos y otros estadísticos útiles para la información de diagnóstico. Cada selección añade una o más variables nuevas a su archivo de datos activo.

Valores pronosticados. Son los valores que el modelo de regresión pronostica para cada caso.

- *Sin estandarizar*. Valor predicho por el modelo para la variable dependiente.
- *Estandarizado*. Transformación de cada valor predicho en su forma estandarizada. Es decir, se sustrae el valor predicho medio al valor predicho y el resultado se divide por la desviación estándar de los valores pronosticados. Los valores pronosticados estandarizados tienen una media de 0 y una desviación estándar de 1.
- *Ajustado*. Valor predicho para un caso cuando dicho caso no se incluye en los cálculos de los coeficientes de regresión.
- *Error estándar de la media*. Errores estándar de los valores pronosticados. Estimación de la desviación estándar del valor promedio de la variable dependiente para los casos que tengan los mismos valores en las variables independientes.

Distancias. Son medidas para identificar casos con combinaciones poco usuales de valores para las variables independientes y casos que puedan tener un gran impacto en el modelo.

- *Mahalanobis*. Medida de cuánto difieren los valores de un caso en las variables independientes del promedio de todos los casos. Una distancia de Mahalanobis grande identifica un caso que tenga valores extremos en una o más de las variables independientes.

- *Cook*. Una medida de cuánto cambiarían los residuos de todos los casos si se excluyera un caso concreto del cálculo de los coeficientes de regresión. Una Distancia de Cook grande indica que la exclusión de ese caso del cálculo de los estadísticos de regresión hará variar substancialmente los coeficientes.
- *Valores de influencia*. Mide la influencia de un punto en el ajuste de la regresión. Influencia centrada varía entre 0 (no influye en el ajuste) a (N-1)/N.

Intervalos de predicción. Los límites superior e inferior para los intervalos de predicción individual y promedio.

- *Media*. Límites inferior y superior (dos variables) para el intervalo de predicción de la respuesta pronosticada promedio.
- *Individual*. Límites superior e inferior (dos variables) del intervalo de predicción para la variable dependiente para un caso individual.
- *Intervalo de confianza*. Especifique un valor entre 1 y 99.99 para especificar el nivel de confianza para los dos intervalos de predicción. Debe seleccionar Media o Individuos antes de introducir este valor. Los valores habituales de los intervalos de confianza son 90, 95 y 99.

Residuos. El valor actual de la variable dependiente menos el valor predicho por la ecuación de regresión.

- *Sin estandarizar*. Diferencia entre un valor observado y el valor predicho por el modelo.
- *Estandarizado*. El residuo dividido por una estimación de su desviación estándar. Los residuos estandarizados, que son conocidos también como los residuos de Pearson o residuos estandarizados, tienen una media de 0 y una desviación estándar de 1.
- *Estudentizado*. El residuo dividido por una estimación de su desviación estándar que varía de un caso a otro, dependiendo de la distancia de los valores de cada caso en las variables independientes de las medias de las variables independientes. A veces se hace referencia a ellos como residuos estudentizados internamente.
- *Suprimido*. El residuo de un caso cuando ese caso se excluye del cálculo de los coeficientes de regresión. Es igual a la diferencia entre el valor de la variable dependiente y el valor predicho corregido.
- *Suprimido estudentizado*. El residuo suprimido para un caso dividido por su error estándar. La diferencia entre un residuo eliminado estudentizado y su residuo estudentizado asociado indica cuánta diferencia hace la eliminación de un caso en su propia predicción. A veces se hace referencia a ellos como residuos estudentizados externamente.

Estadísticos de influencia. El cambio en los coeficientes de regresión (DfBeta) y en los valores pronosticados (DfAjuste) que resulta de la exclusión de un caso particular. También están disponibles los valores tipificados para las DfBeta y para las DfAjuste, junto con la razón entre covarianzas.

- *DfBetas*. La diferencia en el valor beta es el cambio en el coeficiente de regresión que resulta de la exclusión de un caso determinado. Se calcula un valor para cada término del modelo, incluyendo la constante.
- *DfBeta estandarizada*. Diferencia tipificada en el valor beta. El cambio en un coeficiente de regresión cuando se elimina del análisis un caso particular. Puede interesarle examinar aquellos casos cuyos valores absolutos sean mayores que 2 dividido por la raíz cuadrada de N, donde N es el número de casos. Se calcula un valor para cada término del modelo, incluyendo la constante.
- *DfFit*. La diferencia en el valor ajustado es el cambio en el valor predicho que resulta de la exclusión de un caso particular.
- *DfFit estandarizado*. Diferencia tipificada en el valor de ajuste. El cambio en el valor predicho que resulta de la exclusión de un caso particular. Puede interesarle examinar aquellos valores estandarizados cuyo valor absoluto sea mayor que 2 dividido por la raíz cuadrada de p/N, donde p es el número de variables independientes en la ecuación y N es el número de casos.
- *Relación de covarianza*. La razón del determinante de la matriz de covarianzas con un caso particular excluido del cálculo de los coeficientes de regresión respecto al determinante de la matriz de covarianzas con todos los casos incluidos. Si la razón se aproxima a 1, el caso no altera significativamente la matriz de covarianza.

Estadísticos de los coeficientes. Almacena los coeficientes de regresión en un conjunto de datos o en un archivo de datos. Los conjuntos de datos están disponibles para su uso posterior durante la misma sesión, pero no se guardarán como archivos a menos que se hayan guardado explícitamente antes de que finalice la sesión. El nombre de un conjunto de datos debe cumplir las normas de denominación de variables.

Exportar información del modelo a un archivo XML. Las estimaciones de los parámetros y (si lo desea) sus covarianzas se exportan al archivo especificado en formato XML (PMML). Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo.

Regresión lineal: Estadísticos

Se encuentran disponibles los siguientes estadísticos:

Coeficientes de regresión- Estimaciones muestra el coeficiente de regresión *B*, error estándar de *B*, coeficiente beta tipificado, valor *t* para *B*y nivel de significación bilateral de *t*. **Intervalos de confianza** muestra los intervalos de confianza con el nivel de confianza especificado para cada coeficiente de regresión o una matriz de covarianzas. **Matriz de covarianzas** muestra una matriz de varianzas-covarianzas de los coeficientes de regresión, con las covarianzas fuera de la diagonal y las varianzas en la diagonal. También se muestra una matriz de correlaciones.

Ajuste del modelo- Se listan las variables introducidas y eliminadas del modelo y se muestran los siguientes 'estadísticos de bondad de ajuste': varios *R*, R^2 y R^2corregidos, error estándar de la estimación y una tabla de análisis de varianza.

Cambio de R cuadrado- El cambio en el estadístico R^2 que se genera al añadir o suprimir una variable independiente. Si es grande el cambio en R^{cuadrado} asociado a una variable, esto significa que esa variable es un buen predictor de la variable dependiente.

Descriptivo- Proporciona el número de casos válidos, la media y la desviación estándar de cada variable en el análisis. También muestra una matriz de correlaciones con el nivel de significación unilateral y el número de casos para cada correlación.

Correlación de componente. La correlación entre la variable dependiente y una variable independiente cuando los efectos lineales de las otras variables independientes del modelo se han eliminado de la variable independiente. Está relacionada con el cambio en R cuadrado cuando una variable se añade a una ecuación. En ocasiones e denomina correlación semiparcial.

Correlación parcial. La correlación que permanece entre dos variables después de eliminar la correlación que se debe a su asociación mutua con las otras variables. La correlación entre una variable dependiente y una variable independiente cuando se han eliminado de ambas los efectos lineales de las otras variables independientes del modelo.

Diagnóstico de colinealidad- La colinealidad (o multicolinealidad) es la situación no deseable cuando una variable independiente es una función lineal de otras variables independientes. Muestra los autovalores de la matriz de productos vectoriales no centrada y escalada, los índices de condición y las proporciones de la descomposición de la varianza junto con los factores de inflación de la varianza (FIV) y las tolerancias para las variables individuales.

Criterios de selección: incluye el criterio de información de Akaike (AIC), el criterio de predicción de Ameniya (PC), la media cuadrática de error condicional de criterio de predicción (Cp) y el criterio bayesiano de Schwarz (SBC). Los estadísticos se muestran en la tabla de resumen Modelo.

Residuos- Puede seleccionar **'PRESS Statistic'** para utilizarlo como estadístico de validación cruzada para comparar distintos modelos. También muestra la prueba **'Durban-Watson'** para la correlación en serie de los residuos. Elija la información de **'Diagnóstico por casos'** para los casos que cumplen el criterio de selección (valores atípicos por encima de *n* desviaciones estándar).

Regresión lineal: Opciones

Se encuentran disponibles las siguientes opciones:

Criterios del método por pasos. Estas opciones se aplican cuando se especifica el método de selección de variables hacia delante, hacia atrás o por pasos. Las variables se pueden introducir o eliminar del modelo depende de la significación (probabilidad) del valor *F* o del propio valor *F* .

- *Usar probabilidad de F*. Una variable se introduce en el modelo si el nivel de significación de su valor F es menor que el valor de Entrada y se elimina si el nivel de significación es mayor que el valor de Eliminación. La entrada debe ser menor que la eliminación y ambos valores deben ser positivos. Para introducir más variables en el modelo, aumente el valor de entrada. Para eliminar más variables del modelo, disminuya el valor de eliminación.
- *Usar valor de F*. Se introduce una variable en el modelo si su valor F es mayor que el valor de Entrada y se elimina si el valor F es menor que el valor de Eliminación. La entrada debe ser mayor que la eliminación y ambos valores deben ser positivos. Para introducir más variables en el modelo, disminuya el valor de entrada. Para eliminar más variables del modelo, eleve el valor de eliminación.

Tolerancia. De forma predeterminada, el valor es .0001. La tolerancia es la proporción de la varianza de una variable en la ecuación que no está explicada por otras variables independientes en la ecuación. La tolerancia mínima para cualquier variable de la ecuación si la variable considerada fue incluida en el análisis es la tolerancia mínima de una variable que no está incluida en la ecuación. Las variables deben pasar las pruebas de tolerancia y tolerancia mínima para entrar y permanecer en una ecuación de regresión. Si una variable pasa los criterios de tolerancia, es elegible para su inclusión basándose en el método en vigor.

Incluir la constante en la ecuación. De forma predeterminada, el modelo de regresión incluye un término constante. Si se anula la selección de esta opción se obtiene la regresión que pasan por el origen, lo cual se hace raramente. Algunos resultados de la regresión que pasan por el origen no son comparables con los resultados de la regresión que sí incluyen una constante. Por ejemplo, R^2 no puede interpretarse de la manera habitual.

Valores perdidos. Puede elegir uno de los siguientes:

- **Excluir casos según lista.** Sólo se incluirán en el análisis los casos con valores válidos para todas las variables.
- **Excluir casos según pareja.** Los casos con datos completos para la pareja de variables correlacionadas se utilizan para calcular el coeficiente de correlación en el cual se basa el análisis de regresión. Los grados de libertad se basan en el *N* mínimo de las parejas.
- **Reemplazar por la media.** Se emplean todos los casos en los cálculos, sustituyendo las observaciones perdidas por la media de la variable.

Características adicionales del comando REGRESSION

La sintaxis de comandos también le permite:

- Escribir una matriz de correlaciones o leer una matriz (en lugar de los datos en bruto) con el fin de obtener el análisis de regresión (mediante el subcomando `MATRIX`).
- Especificar los niveles de tolerancia (mediante el subcomando `CRITERIA`).
- Obtener múltiples modelos para las mismas variables dependientes u otras diferentes (mediante los subcomandos `METHOD` y `DEPENDENT`).
- Obtener estadísticas adicionales (con los subcomandos `DESCRIPTIVES` y `STATISTICS`).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Regresión ordinal

La regresión ordinal permite dar forma a la dependencia de una respuesta ordinal politómica sobre un conjunto de predictores, que pueden ser factores o covariables. El diseño de la regresión ordinal se basa en la metodología de McCullagh (1980, 1998) y en la sintaxis se hace referencia al procedimiento como `PLUM`.

El análisis de regresión lineal ordinario implica minimizar las diferencias de la suma de los cuadrados entre una variable de respuesta (la dependiente) y una combinación ponderada de las variables predictoras (las independientes). Los coeficientes estimados reflejan cómo los cambios en los predictores afectan a la respuesta. Se considera que la respuesta es numérica, en el sentido en que los cambios en el nivel de la respuesta son equivalentes en todo el rango de la respuesta. Por ejemplo, la diferencia de altura entre una persona que mide 150 cm y una que mide 140 cm es de 10 cm, que tiene el mismo significado que la diferencia de altura entre una persona que mide 210 cm y una que mide 200 cm. Estas relaciones no se mantienen necesariamente con las variables ordinales, en las que la elección y el número de categorías de respuesta pueden ser bastante arbitrarios.

Ejemplo. La regresión ordinal podría utilizarse para estudiar la reacción de los pacientes con respecto a una dosis de un fármaco. Las reacciones posibles podrían clasificarse como *ninguna*, *ligera*, *moderada* o *grave*. La diferencia entre una reacción ligera y una moderada es difícil o imposible de cuantificar y se basa en la apreciación. Además, la diferencia entre una respuesta ligera y una moderada podría ser superior o inferior a la diferencia entre una respuesta moderada y una grave.

Estadísticos y gráficos. Frecuencias observadas y esperadas y frecuencias acumuladas, residuos de Pearson para las frecuencias y las frecuencias acumuladas, probabilidades observadas y esperadas, probabilidades acumuladas observadas y esperadas para cada categoría de respuesta por patrón en las covariables, matrices de correlaciones asintóticas y de covarianzas entre las estimaciones de los parámetros, chi-cuadrado de Pearson y chi-cuadrado de la razón de verosimilitud, estadísticos de bondad de ajuste, historial de iteraciones, contraste del supuesto de líneas paralelas, estimaciones de los parámetros, errores estándar, intervalos de confianza y estadísticos R^2 de Cox y Snell, de Nagelkerke y de McFadden.

Regresión ordinal: Consideraciones sobre los datos

Datos. Se asume que la variable dependiente es ordinal y puede ser numérica o de cadena. El orden se determina al clasificar los valores de la variable dependiente en orden ascendente. El valor inferior define la primera categoría. Se asume que las variables de factor son categóricas. Las covariables deben ser numéricas. Observe que al usar más de una covariable continua, se puede llegar a crear una tabla de probabilidades de casilla muy grande.

Supuestos. Sólo se permite una variable de respuesta y debe especificarse. Además, para cada patrón distinto de valores en las variables independientes, se supone que las respuestas son variables multinomiales independientes.

Procedimientos relacionados. La regresión logística nominal utiliza modelos similares para las variables dependientes nominales.

Obtener una regresión ordinal

1. En los menús seleccione:

   Analizar > **Regresión** > **Ordinal...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una variable dependiente.
3. Pulse **Aceptar**.

Regresión ordinal: Opciones

El cuadro de diálogo Opciones le permite ajustar los parámetros utilizados en el algoritmo de estimación iterativo, seleccionar un nivel de confianza para las estimaciones de los parámetros y seleccionar una función de enlace.

Iteraciones. El algoritmo iterativo puede personalizarse.

- **Número máximo de iteraciones.** Especifique un número entero no negativo. Si se especifica el 0, el procedimiento devolverá las estimaciones iniciales.
- **Máxima subdivisión por pasos.** Especifique un número entero positivo.

- **Convergencia del logaritmo de la verosimilitud.** El algoritmo se detiene si el cambio absoluto o relativo en el logaritmo de la verosimilitud es inferior a este valor. Si se especifica 0, no se utiliza el criterio.
- **Convergencia de los parámetros.** El algoritmo se detiene si el cambio absoluto o relativo en cada una de las estimaciones de los parámetros es inferior a este valor. Si se especifica 0, no se utiliza el criterio.

Intervalo de confianza. Especifique un valor mayor o igual a 0 e inferior a 100.

Delta. El valor añadido a las frecuencias de casilla de cero. Especifique un valor no negativo inferior a 1.

Tolerancia para la singularidad. Utilizada para comprobar los predictores con alta dependencia. Seleccione un valor en la lista de opciones.

Función de enlace. La función de enlace es una transformación de las probabilidades acumuladas que permiten la estimación del modelo. Se encuentran disponibles las cinco funciones de enlace siguientes.

- **Logit.** $f(x)=\log(x/(1-x))$. Se utiliza normalmente para categorías distribuidas uniformemente.
- **Log-log complementario.** $f(x)=\log(-\log(1-x))$. Se utiliza normalmente cuando las categorías más altas son más probables.
- **Log-log negativo.** $f(x)=-\log(-\log(x))$. Se utiliza normalmente cuando las categorías más bajas son más probables.
- **Probit.** $f(x)=\Phi^{-1}(x)$. Se utiliza normalmente cuando la variable latente sigue una distribución normal.
- **Cauchit (Cauchy inversa).** $f(x)=\tan(\pi(x-0.5))$. Se utiliza normalmente cuando la variable latente tiene muchos valores extremos.

Resultado de la regresión ordinal

El cuadro de diálogo Resultados le permite generar tablas que se pueden visualizar en el Visor y guardar variables en el archivo de trabajo.

Representación. Genera tablas correspondientes a:

- **Imprimir el historial de iteraciones para cada n pasos.** El logaritmo de la verosimilitud y las estimaciones de los parámetros se imprimen con la frecuencia de iteraciones a imprimir especificada. Siempre se imprimen la primera y la última iteración.
- **Estadísticos de bondad de ajuste.** Estadísticos chi-cuadrado de Pearson y chi-cuadrado de la razón de verosimilitud. Estos estadísticos se calculan según la clasificación especificada en la lista de variables.
- **Estadísticos de resumen.** R^2 de Cox y Snell, de Nagelkerke y de McFadden.
- **Estimaciones de los parámetros.** Estimaciones de los parámetros, errores estándar e intervalos de confianza.
- **Correlación asintótica de las estimaciones de los parámetros.** Matriz de las correlaciones entre las estimaciones de los parámetros.
- **Covarianza asintótica de las estimaciones de los parámetros.** Matriz de las covarianzas entre las estimaciones de los parámetros.
- **Información de casilla.** Frecuencias esperadas y observadas y frecuencias acumuladas, residuos de Pearson para las frecuencias y las frecuencias acumuladas, probabilidades esperadas y observadas y probabilidades esperadas y observadas de cada categoría de respuesta según el patrón en las covariables. Tenga en cuenta que en el caso de modelos con muchos patrones de covariables (por ejemplo, modelos con covariables continuas), esta opción puede generar una tabla grande y poco manejable.
- **Prueba de líneas paralelas.** Prueba correspondiente a la hipótesis de que los parámetros de ubicación son equivalentes en todos los niveles de la variable dependiente. Esta prueba está disponible únicamente para el modelo de sólo ubicación.

Variables guardadas. Guarda las siguientes variables en el archivo de trabajo:

- **Probabilidades de respuesta estimadas.** Probabilidades estimadas por el modelo para la clasificación de un patrón de factores/covariables en las categorías de respuesta. El número de probabilidades es igual al número de categorías de respuesta.
- **Categoría pronosticada.** La categoría de respuesta con la mayor probabilidad estimada para un patrón de factores/covariables.
- **Probabilidad de la categoría pronosticada.** Probabilidades estimada de la clasificación de un patrón factores/covariables en la categoría pronosticada. Esta probabilidad también es el máximo de las probabilidades estimadas para el patrón de factores/covariables.
- **Probabilidad de la categoría real.** Probabilidad estimada de la clasificación de un patrón de factores/covariables en la categoría real.

Imprimir log-verosimilitud. Controla la representación del logaritmo de la verosimilitud. **Incluir constante multinomial** ofrece el valor completo de la verosimilitud. Para comparar los resultados con los productos que no incluyan la constante, puede seleccionar la opción de excluirla.

Modelo de ubicación de la regresión ordinal

El cuadro de diálogo Ubicación le permite especificar el modelo de ubicación para el análisis.

Especificar modelo. Un modelo de efectos principales contiene los efectos principales de las covariables y los factores, pero no contiene efectos de interacción. Puede crear un modelo personalizado para especificar subconjuntos de interacciones entre los factores o bien interacciones entre las covariables.

Factores y covariables. Muestra una lista de los factores y las covariables.

Modelo de ubicación. El modelo depende de los efectos principales y de los de interacción que seleccione.

Para las covariables y los factores seleccionados:

Interacción
Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el valor predeterminado.

Efectos principales
Crea un término de efectos principales para cada variable seleccionada.

Todas de 2
Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todas de 3
Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todas de 4
Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todas de 5
Crea todas las interacciones quíntuples posibles de las variables seleccionadas.

Construcción de términos y Términos personalizados

Crear términos
Utilice esta opción si desea incluir términos no anidados de un tipo determinado (por ejemplo, efectos principales) para todas las combinaciones de un conjunto seleccionado de factores y covariables.

Construir términos personalizados
Utilice esta opción si desea incluir términos anidados o si desea crear explícitamente una variable de término por variable. La creación de un término anidado implica los pasos siguientes:

Modelo de escala de la regresión ordinal

El cuadro de diálogo Escala le permite especificar el modelo de escala para el análisis.

Factores y covariables. Muestra una lista de los factores y las covariables.

Modelo de escala. El modelo depende de los efectos principales y de los de interacción que seleccione.

Para las covariables y los factores seleccionados:

Interacción
Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el valor predeterminado.

Efectos principales
Crea un término de efectos principales para cada variable seleccionada.

Todas de 2
Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todas de 3
Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todas de 4
Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todas de 5
Crea todas las interacciones quíntuples posibles de las variables seleccionadas.

Construcción de términos y Términos personalizados

Crear términos
Utilice esta opción si desea incluir términos no anidados de un tipo determinado (por ejemplo, efectos principales) para todas las combinaciones de un conjunto seleccionado de factores y covariables.

Construir términos personalizados
Utilice esta opción si desea incluir términos anidados o si desea crear explícitamente una variable de término por variable. La creación de un término anidado implica los pasos siguientes:

Características adicionales del comando PLUM

Se puede personalizar la regresión ordinal si se pegan las selecciones en una ventana de sintaxis y se edita la sintaxis del comando PLUM resultante. La sintaxis de comandos también le permite:

- Crear contrastes de hipótesis personalizados especificando las hipótesis nulas como combinaciones lineales de los parámetros.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Regresión lineal de Elastic Net

Linear Elastic Net utiliza la clase Python sklearn.linear_model.ElasticNet para estimar los modelos de regresión lineal regularizada para una variable dependiente en una o más variables independientes. La regularización combina las penalizaciones L1 (Lasso) y L2 (Ridge). La extensión incluye modalidades opcionales para visualizar gráficos de rastreo para distintos valores de alfa para una proporción L1 determinada, y para seleccionar la proporción L1 y los valores de hiperparámetro alfa basados en la validación cruzada. Cuando se ajusta un único modelo o se utiliza la validación cruzada para seleccionar la relación de penalización y/o alfa, se puede utilizar una partición de datos reservados para estimar el rendimiento fuera de la muestra.

Además de ajustar un modelo con valores especificados de la razón de la penalización L1 y el parámetro de regularización alfa, la red elástica lineal puede mostrar un trazo de valores de coeficiente para un rango de valores alfa para una razón determinada, o facilitar la elección del valor de hiperparámetros a través de la validación cruzada k-fold en cuadrículas de valores especificadas. Si se ajusta un modelo unico o se realiza la seleccion de proporcion y/o alfa a traves de la validacion cruzada, el modelo final se puede aplicar a los datos retenidos que se crean por una division de los datos de entrada para obtener una estimacion valida del desempeno fuera de la muestra del modelo.

Obtención de un análisis de regresión de red elástica lineal

1. En los menús seleccione:

   Analizar > **Regresión** > **Alternativas OLS lineales** > **Red elástica**

   El diálogo le permite especificar una variable que asigna cada caso del conjunto de datos activo a la muestra de entrenamiento o reservada.

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una variable objetivo numérica. Sólo se necesita una variable objetivo para ejecutar un análisis.
3. Especifique un dependiente numérico.
4. Especifique al menos una variable de factor categórico o una variable de covariable numérica.

Opcionalmente, **Partición** proporciona una forma de crear una reserva o un subconjunto de prueba de los datos de entrada para la estimación del rendimiento fuera de muestra del modelo especificado o elegido. Todo el particionamiento se realiza después de la supresión por lista de los casos con datos no válidos para cualquier variable utilizada por el procedimiento. Tenga en cuenta que para la validación cruzada, se crean pliegues o particiones de los datos de entrenamiento en Python. Los datos reservados creados por la partición no se utilizan en la estimación, independientemente de la modalidad en vigor.

La partición se puede definir especificando la proporción de casos que se asignan aleatoriamente a cada muestra (bajo **Particiones de entrenamiento y retención**), o mediante una variable que asigna cada caso a la muestra de entrenamiento o reserva. No puede especificar tanto el entrenamiento como las variables. Si no se especifica la partición, se crea una muestra reservada de aproximadamente el 30% de los datos de entrada.

El **% de entrenamiento** especifica el número relativo de casos en el conjunto de datos activo que se asignará aleatoriamente a la muestra de entrenamiento. El entrenamiento predeterminado es 70%.

Regresión de red elástica lineal: Opciones

La pestaña Opciones proporciona opciones para:

Modo

Esta selección proporciona opciones para especificar una de las modalidades siguientes:

Ajustar con alfa y la proporción de L1 especificada
Cuando selecciona esta modalidad, se ajusta un único modelo a los datos de entrenamiento utilizando la proporción L1 y los valores de regularización alfa especificados. Este es el valor predeterminado. Si se especifica una partición, el modelo único o final que se ajusta se aplica a los datos de prueba retenidos para estimar el rendimiento fuera de muestra.

En **Gráfico**, se pueden seleccionar gráficos de los valores observados y/o de los residuos frente a los valores pronosticados.

En **Guardar**, puede especificar los valores pronosticados y los residuos que desea guardar.

Gráfico de rastreo
Cuando selecciona esta modalidad, se muestran tres gráficos para los datos de entrenamiento como una función de alfa para el conjunto especificado de valores de alfa:

- Un gráfico de rastreo de coeficientes de regresión.
- Un gráfico de R^2.
- Gráfico de error cuadrático medio (MSE).

Aunque se respeta la partición, no se proporcionan resultados para los datos de prueba retenidos porque no hay resultados de modelo finales de esta modalidad.

L1 proporción y/o selección alfa mediante validación cruzada
Cuando selecciona esta modalidad, se realiza una búsqueda de cuadrícula con validación cruzada para evaluar modelos, y se eligen los valores de mejor proporción y alfa basándose en el mejor promedio de R^2 en los pliegues de validación. El campo **Número de pliegues de validación cruzada** se puede utilizar para cambiar el valor predeterminado de cinco divisiones o pliegues para la validación cruzada. Si se especifica una partición, el modelo único o final que se ajusta se aplica a los datos de prueba retenidos para estimar el rendimiento fuera de muestra.

En **Visualizar**, puede seleccionar mostrar información básica sólo sobre el modelo con el valor elegido de razón y alfa (**Mejor**), información básica sobre todos los modelos comparados (**Comparar modelos**) o información completa sobre todas las divisiones o pliegues para todos los modelos (**Comparar modelos y pliegues**). **Mejor** es el valor predeterminado.

En **Gráfico**, se pueden seleccionar gráficos de la media R^2 y/o MSE sobre pliegues de validación. También se pueden seleccionar los gráficos de los valores observados y/o los residuos frente a los valores pronosticados.

En **Guardar**, puede especificar los valores pronosticados y los residuos que desea guardar.

Especificar proporciones de L1 individuales
Cuando selecciona esta modalidad para la modalidad **Ajustar con la proporción L1 especificada y la modalidad alfa** o **Rastrear gráfico** , puede especificar un único valor de proporción de penalización L1 . Cuando se selecciona para la modalidad de **L1 proporción y/o selección alfa mediante validación cruzada** , puede especificar varios valores.

Especificar cuadrícula de valores alfa
Cuando selecciona esta modalidad para la modalidad de **L1 proporción y/o selección alfa mediante validación cruzada** , se puede especificar una cuadrícula de valores alfa exclusivos desde un valor de **Inicio** (`value1`) a un valor de **Fin** (`value2`) con el incremento de **Por** (`value3`). Si se especifica, sólo se permite un conjunto válido de [`value1 TO value2 BY value3`]. Debe satisfacer ese $0 \leq value1 \leq value2 \leq 1$. En los casos en los que $value1 = value2$, es equivalente a especificar un único `value1`, independientemente de `value3`.

Los gráficos se muestran utilizando la métrica especificada para los ejes X horizontales de los valores alfa variables.

Especificar valores alfa individuales
Cuando selecciona esta modalidad para la modalidad **Ajustar con la proporción L1 y alfa** especificadas, puede especificar un único valor de regularización de alfa. Cuando se selecciona para la modalidad **Trazar gráfico** o **L1 de proporción y/o selección alfa mediante validación cruzada** , puede especificar varios valores.

Especificar cuadrícula de valores alfa
Cuando selecciona esta modalidad para la modalidad **Trazar gráfico** o **L1 de proporción y/o selección alfa mediante validación cruzada** , se puede especificar una cuadrícula de valores alfa exclusivos desde un valor de **Inicio** (`value1`) a un valor de **Fin** (`value2`) con el incremento de **Por** (`value3`). Si se especifica, sólo se permite un conjunto válido de [`value1 TO value2 BY value3`]. Debe satisfacer ese $0 \leq value1 \leq value2 \leq 1$. En los casos en los que $value1 = value2$, es equivalente a especificar un único `value1`, independientemente de `value3`.

La **métrica Alfa** para rangos de valores puede ser **Lineal** o **Base 10 logarítmica** (10 se eleva a la potencia de los valores especificados).

Los gráficos se muestran utilizando la métrica especificada para los ejes X horizontales de los valores alfa variables.

Criterios

Controla los análisis.

Incluir intersección
Este criterio incluye una interceptación en los modelos ajustados. Tenga en cuenta que el procedimiento de extensión no centra ni estandariza la variable dependiente, y la intersección no se penaliza durante la estimación.

Estandarizar predictores
Estandariza todas las variables independientes.

Número de pliegues de validación cruzada
Número de divisiones o pliegues para la evaluación de validación cruzada de modelos. Debe ser un valor entero positivo mayor que 1. El valor predeterminado es 5.

Estado aleatorio de Python
El valor de random_state en Python utilizado al realizar la evaluación de validación cruzada de modelos. Permite la réplica de resultados que implican números pseudo-aleatorios. El valor debe ser un entero en el rango de 0 a 2^{32}-1. El valor predeterminado es 0.

Límite de tiempo (minutos)
El número de minutos permitidos para que se ejecuten los cálculos del modelo. Si especifica 0, el temporizador se desactiva. El valor predeterminado es 5.

Visualización

Esta opción especifica la cantidad de salida que se va a visualizar para la proporción **L1 y/o la selección alfa mediante la modalidad de validación cruzada** .

El mejor
Muestra sólo los resultados básicos para el mejor modelo elegido. Se establece de forma predeterminada.

Comparar modelos
Muestra los resultados básicos para todos los modelos evaluados.

Comparar modelos y pliegues
Muestra resultados detallados completos para cada división o plegado para cada modelo evaluado.

Gráfico

Esta opción especifica gráficos de valores observados o residuales frente a valores pronosticados, y con validación cruzada, especificación de gráficos de promedio de error cuadrático promedio (MSE) y/o promedio R^2 sobre pliegues de validación cruzada frente a valores alfa.

Promedio de error cuadrático promedio de validación cruzada (MSE) frente a alfa
Para la modalidad de **L1 proporción y/o selección alfa mediante validación cruzada** , muestra un gráfico de líneas de promedio de MSE sobre pliegues de validación cruzada frente a alfa para el mejor valor de proporción L1 especificado o seleccionado. Para la modalidad de **Gráfico de rastreo** , se genera automáticamente un gráfico similar basándose en los datos de entrenamiento completos.

Promedio de R cuadrado de validación cruzada frente a alfa
Para la modalidad de **L1 proporción y/o selección alfa mediante validación cruzada** , muestra un gráfico de líneas de R^2 promedio sobre pliegues de validación cruzada frente a alfa para el mejor valor de proporción L1 especificado o seleccionado. Para la modalidad de **Gráfico de rastreo** , se genera automáticamente un gráfico similar basándose en los datos de entrenamiento completos.

Observado frente a Predicho
Muestra un diagrama de dispersión de los valores observados frente a los previstos para el modelo especificado o mejor.

Residuos frente a pronosticados
Muestra un diagrama de dispersión de los residuos frente a los valores pronosticados para el modelo especificado o mejor.

Guardar

Especifica variables para guardar en el conjunto de datos activo.

Valores pronosticados
Guarde los valores pronosticados del modelo especificado o mejor en el conjunto de datos activo. También puede especificar un **Nombre de variable personalizada**.

Residuos
Guarde los residuos de las predicciones de modelo especificadas o mejores en el conjunto de datos activo. También puede especificar un **Nombre de variable personalizada**.

Regresión lineal de Lasso

Lasso lineal utiliza la clase Python sklearn.linear_model.Lasso para estimar los modelos de regresión lineal regularizada por pérdida L1 para una variable dependiente en una o más variables independientes, e incluye modalidades opcionales para visualizar gráficos de rastreo y para seleccionar el valor de hiperparámetro alfa basado en la validación cruzada. Cuando se ajusta un único modelo o se utiliza la validación cruzada para seleccionar alfa, se puede utilizar una partición de datos reservados para estimar el rendimiento fuera de la muestra.

Además de ajustar un modelo con un valor especificado del parámetro de regularización alfa, el lazo lineal puede mostrar un trazo de valores de coeficiente para un rango de valores alfa, o facilitar la elección del valor de hiperparámetro a través de la validación cruzada de k-fold en cuadrículas especificadas de valores. Si se ajusta un único modelo o se realiza una selección alfa mediante validación cruzada, el modelo final se puede aplicar a los datos retenidos que se crean mediante una partición de los datos de entrada para obtener una estimación válida del rendimiento fuera de la muestra del modelo.

Obtención de un análisis de regresión de lazo lineal

1. En los menús seleccione:

   Analizar > **Regresión** > **Alternativas OLS lineales** > **Lasso**

   El diálogo le permite especificar una variable que asigna cada caso del conjunto de datos activo a la muestra de entrenamiento o reservada.

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una variable objetivo numérica. Sólo se necesita una variable objetivo para ejecutar un análisis.
3. Especifique un dependiente numérico.
4. Especifique al menos una variable de factor categórico o una variable de covariable numérica.

Opcionalmente, **Partición** proporciona una forma de crear una reserva o un subconjunto de prueba de los datos de entrada para la estimación del rendimiento fuera de muestra del modelo especificado o elegido. Todo el particionamiento se realiza después de la supresión por lista de los casos con datos no válidos para cualquier variable utilizada por el procedimiento. Tenga en cuenta que para la validación cruzada, se crean pliegues o particiones de los datos de entrenamiento en Python. Los datos reservados creados por la partición no se utilizan en la estimación, independientemente de la modalidad en vigor.

La partición se puede definir especificando la proporción de casos asignados aleatoriamente a cada muestra (bajo **Particiones de entrenamiento y retención**), o mediante una variable que asigna cada caso a la muestra de entrenamiento o reserva. No puede especificar tanto el entrenamiento como las variables. Si no se especifica la partición, se crea una muestra reservada de aproximadamente el 30% de los datos de entrada.

El **% de entrenamiento** especifica el número relativo de casos en el conjunto de datos activo que se asignará aleatoriamente a la muestra de entrenamiento. El entrenamiento predeterminado es 70%.

Regresión de lazo lineal: Opciones

La pestaña Opciones proporciona opciones para:

Modo

Esta selección proporciona opciones para especificar una de las modalidades siguientes:

Ajustar con alfa especificado
Cuando selecciona esta modalidad, se ajusta un único modelo a los datos de entrenamiento utilizando sólo un valor de regularización alfa. Se establece de forma predeterminada. Si se especifica una partición, el modelo único o final que se ajusta se aplica a los datos de prueba retenidos para estimar el rendimiento fuera de muestra.

En **Gráfico**, se pueden seleccionar gráficos de los valores observados y/o de los residuos frente a los valores pronosticados.

En **Guardar**, puede especificar los valores pronosticados y los residuos que desea guardar.

Gráfico de rastreo
Cuando selecciona esta modalidad, se muestran tres gráficos para los datos de entrenamiento como una función de alfa para el conjunto especificado de valores de alfa:

- Un gráfico de rastreo de coeficientes de regresión.
- Un gráfico de R^2.
- Gráfico de error cuadrático medio (MSE).

Aunque se respeta la partición, no se proporcionan resultados para los datos de prueba retenidos porque no hay resultados de modelo finales de esta modalidad.

Selección alfa mediante validación cruzada

Cuando selecciona una búsqueda de cuadrícula con validación cruzada para evaluar modelos, y elige el mejor alfa basándose en el mejor promedio, R^2 sobre los pliegues de validación. El campo **Número de pliegues de validación cruzada** se puede utilizar para cambiar el valor predeterminado de cinco divisiones o pliegues para la validación cruzada. Si se especifica una partición, el modelo único o final que se ajusta se aplica a los datos de prueba retenidos para estimar el rendimiento fuera de muestra.

En **Visualización**, puede seleccionar mostrar información básica sólo sobre el modelo con el valor elegido de alfa (**Mejor**), información básica sobre todos los modelos comparados (**Comparar modelos**) o información completa sobre todas las divisiones o pliegues para todos los modelos (**Comparar modelos y pliegues**). **Mejor** es el valor predeterminado.

En **Gráfico**, se pueden seleccionar gráficos de la media R^2 y/o MSE sobre pliegues de validación. También se pueden seleccionar los gráficos de los valores observados y/o los residuos frente a los valores pronosticados.

En **Guardar**, puede especificar los valores pronosticados y los residuos que desea guardar.

Especificar valores alfa individuales
Cuando selecciona la modalidad **Ajustar con alfa especificado** , puede especificar un único valor de regularización de alfa. Cuando selecciona el modo **Rastrear gráfico** o **Selección alfa mediante validación cruzada** , puede especificar varios valores.

Valores
Especifique uno o varios valores de regularización alfa positivos. Se pueden especificar varios valores individualmente o como rangos. El valor predeterminado es 1.

Especificar cuadrícula de valores alfa
Cuando selecciona la modalidad **Rastrear gráfico** o **Selección alfa mediante validación cruzada** , se puede especificar una cuadrícula de valores alfa exclusivos desde un valor de **Inicio** (`value1`) a un valor de **Fin** (`value2`) con el incremento de **Por** (`value3`). Si se especifica, sólo se permite un conjunto válido de [`value1 TO value2 BY value3`]. Debe satisfacer ese `0 ≤ value1 ≤ value2 ≤ 1`. En los casos en los que `value1 = value2`, es equivalente a especificar un único `value1`, independientemente de `value3`.

La **métrica Alfa** para rangos de valores puede ser **Lineal** o **Base 10 logarítmica** (10 se eleva a la potencia de los valores especificados).

Los gráficos se muestran utilizando la métrica especificada para los ejes X horizontales de los valores alfa variables.

Criterios

Controla los análisis.

Incluir intersección
Incluye una interceptación en los modelos ajustados. Tenga en cuenta que el procedimiento de extensión no centra ni estandariza la variable dependiente, y la intersección no se penaliza durante la estimación.

Estandarizar predictores
Estandariza todas las variables independientes.

Número de pliegues de validación cruzada
Número de divisiones o pliegues para la evaluación de validación cruzada de modelos. Debe ser un valor entero positivo mayor que 1. El valor predeterminado es 5.

Estado aleatorio de Python
El valor de random_state en Python utilizado al realizar la evaluación de validación cruzada de modelos. Permite la réplica de resultados que implican números pseudoaleatorios. Debe ser un entero en el rango de 0 a 2^{32}-1. El valor predeterminado es 0.

Límite de tiempo (minutos)
El número de minutos permitidos para que se ejecuten los cálculos del modelo. Si especifica 0, el temporizador se desactiva. El valor predeterminado es 5.

Visualización

Especifica la cantidad de salida que se va a visualizar para la modalidad **Selección alfa mediante validación cruzada** .

El mejor
Muestra sólo los resultados básicos para el mejor modelo elegido. Este es el valor predeterminado.

Comparar modelos
Muestra los resultados básicos para todos los modelos evaluados.

Comparar modelos y pliegues
Muestra resultados detallados completos para cada división o plegado para cada modelo evaluado.

Gráfico

Especifica gráficos de valores observados o residuales frente a valores pronosticados, y con validación cruzada, especificación de gráficos de error cuadrático promedio (MSE) y/o R promedio^2 sobre pliegues de validación cruzada frente a valores alfa.

Promedio de error cuadrático promedio de validación cruzada (MSE) frente a alfa
Para la modalidad **Selección alfa mediante validación cruzada** , muestra un gráfico de líneas de promedio de MSE sobre pliegues de validación cruzada frente a alfa. Para la modalidad de **Gráfico de rastreo** , se genera automáticamente un gráfico similar basándose en los datos de entrenamiento completos.

Promedio de R cuadrado de validación cruzada frente a alfa
Para la modalidad **Selección alfa mediante validación cruzada** , muestra un gráfico de líneas de R^2 promedio sobre pliegues de validación cruzada frente a alfa. Para la modalidad de **Gráfico de rastreo** , se genera automáticamente un gráfico similar basándose en los datos de entrenamiento completos.

Observado frente a Predicho
Muestra un diagrama de dispersión de los valores observados frente a los previstos para el modelo especificado o mejor.

Residuos frente a pronosticados
Muestra un diagrama de dispersión de los residuos frente a los valores pronosticados para el modelo especificado o mejor.

Guardar

Especifica variables para guardar en el conjunto de datos activo.

Valores pronosticados
Guarde los valores pronosticados del modelo especificado o mejor en el conjunto de datos activo. También puede especificar un **Nombre de variable personalizada**.

Residuos
Guarde los residuos de las predicciones de modelo especificadas o mejores en el conjunto de datos activo. También puede especificar un **Nombre de variable personalizada**.

Regresión lineal Ridge

Linear Ridge utiliza la clase Python sklearn.linear_model.Ridge para estimar L2 o los modelos de regresión lineal regularizada por pérdida cuadrada para una variable dependiente en una o más variables independientes, e incluye modalidades opcionales para visualizar gráficos de rastreo y para seleccionar el valor de hiperparámetro alfa basado en la validación cruzada. Cuando se ajusta un único modelo o se utiliza la validación cruzada para seleccionar alfa, se puede utilizar una partición de datos reservados para estimar el rendimiento fuera de la muestra.

Ademas de ajustar un modelo con un valor especificado del parametro de regularizacion alfa, el reborde lineal puede presentar un trazado de reborde de valores de coeficiente para un rango de valores alfa, o facilitar la eleccion del valor de hiperparametro a traves de la validacion cruzada de k-fold en cuadrículas especificadas de valores. Si se ajusta un único modelo o se realiza una selección alfa mediante validación cruzada, el modelo final se puede aplicar a los datos retenidos que se crean mediante una partición de los datos de entrada para obtener una estimación válida del rendimiento fuera de la muestra del modelo.

Obtención de un análisis de regresión de cresta lineal

1. En los menús seleccione:

   Analizar > **Regresión** > **Alternativas de OLS lineal** > **Ridge**

   El diálogo le permite especificar una variable que asigna cada caso del conjunto de datos activo a la muestra de entrenamiento o reservada.
2. Seleccione una variable objetivo numérica. Sólo se necesita una variable objetivo para ejecutar un análisis.
3. Especifique un dependiente numérico.
4. Especifique al menos una variable de factor categórico o una variable de covariable numérica.

Opcionalmente, **Partición** proporciona una forma de crear una reserva o un subconjunto de prueba de los datos de entrada para la estimación del rendimiento fuera de muestra del modelo especificado o elegido. Todo el particionamiento se realiza después de la supresión por lista de los casos con datos no válidos para cualquier variable utilizada por el procedimiento. Tenga en cuenta que para la validación cruzada, se crean pliegues o particiones de los datos de entrenamiento en Python. Los datos reservados creados por la partición no se utilizan en la estimación, independientemente de la modalidad en vigor.

La partición se puede definir especificando la proporción de casos que se asignan aleatoriamente a cada muestra (bajo **Particiones de entrenamiento y retención**), o mediante una variable que asigna cada caso a la muestra de entrenamiento o reserva. No puede especificar tanto el entrenamiento como las variables. Si no se especifica la partición, se crea una muestra reservada de aproximadamente el 30% de los datos de entrada.

El **% de entrenamiento** especifica el número relativo de casos en el conjunto de datos activo que se asignará aleatoriamente a la muestra de entrenamiento. El entrenamiento predeterminado es 70%.

Regresión de cresta lineal: Opciones

La pestaña Opciones proporciona opciones para:

Modo

Esta selección proporciona opciones para especificar una de las modalidades siguientes:

Ajustar con alfa especificado
Cuando selecciona esta opción, se ajusta un único modelo a los datos de entrenamiento que utiliza sólo un valor de regularización alfa. Se establece de forma predeterminada. Si se especifica una partición, el modelo único o final que se ajusta se aplica a los datos de prueba retenidos para estimar el rendimiento fuera de muestra.

En **Gráfico**, se pueden seleccionar gráficos de los valores observados y/o de los residuos frente a los valores pronosticados.

En **Guardar**, puede especificar los valores pronosticados y los residuos que desea guardar.

Gráfico de rastreo
Cuando selecciona esta opción, se muestran tres gráficos para los datos de entrenamiento como una función de alfa para el conjunto especificado de valores de alfa:

- Un gráfico de trazado de cresta de coeficientes de regresión.
- Un gráfico de R^2.
- Gráfico de error cuadrático medio (MSE).

Aunque se respeta la partición, no se proporcionan resultados para los datos de prueba retenidos porque no hay resultados de modelo finales de esta modalidad.

Selección alfa mediante validación cruzada

Cuando selecciona esta opción, se realiza una búsqueda de cuadrícula con validación cruzada para evaluar modelos, y se elige el mejor alfa basándose en el mejor promedio de R^2 en los pliegues de validación. El campo **Número de pliegues de validación cruzada** se puede utilizar para cambiar el valor predeterminado de cinco divisiones o pliegues para la validación cruzada. Si se especifica una partición, el modelo único o final que se ajusta se aplica a los datos de prueba retenidos para estimar el rendimiento fuera de muestra.

En **Visualización**, puede seleccionar mostrar información básica sólo sobre el modelo con el valor elegido de alfa (**Mejor**), información básica sobre todos los modelos comparados (**Comparar modelos**) o información completa sobre todas las divisiones o pliegues para todos los modelos (**Comparar modelos y pliegues**). **Mejor** es el valor predeterminado.

En **Gráfico**, se pueden seleccionar gráficos de la media R^2 y/o MSE sobre pliegues de validación. También se pueden seleccionar los gráficos de los valores observados y/o los residuos frente a los valores pronosticados.

En **Guardar**, puede especificar los valores pronosticados y los residuos que desea guardar.

Especificar valores alfa individuales
Cuando selecciona la modalidad **Ajustar con alfa especificado** , puede especificar un único valor de regularización de alfa. Cuando selecciona el modo **Rastrear gráfico** o **Selección alfa mediante validación cruzada** , puede especificar varios valores.

Valores
Especifique uno o varios valores de regularización alfa positivos. Puede especificar varios valores individualmente o como rangos. El valor predeterminado es 1.

Especificar cuadrícula de valores alfa
Cuando selecciona la modalidad **Rastrear gráfico** o **Selección alfa mediante validación cruzada** , se puede especificar una cuadrícula de valores alfa exclusivos desde un valor de **Inicio** (`value1`) a un valor de **Fin** (`value2`) con el incremento de **Por** (`value3`). Si se especifica, sólo se permite un conjunto válido de [`value1 TO value2 BY value3`]. Debe satisfacer ese `0 ≤ value1 ≤ value2`

≤ 1. En los casos en los que `value1 = value2`, es equivalente a especificar un único `value1`, independientemente de `value3`.

La **métrica Alfa** para rangos de valores puede ser **Lineal** o **Base 10 logarítmica** (10 se eleva a la potencia de los valores especificados).

Los gráficos se muestran utilizando la métrica especificada para los ejes X horizontales de los valores alfa variables.

Criterios

Controla los análisis.

Incluir intersección
Este criterio incluye una interceptación en uno o más modelos ajustados. Tenga en cuenta que el procedimiento de extensión no centra ni estandariza la variable dependiente, y la intersección no se penaliza durante la estimación.

Estandarizar predictores
Este criterio estandariza todas las variables independientes.

Número de pliegues de validación cruzada
Utilice este criterio para establecer el número de divisiones o pliegues para la evaluación de validación cruzada de modelos. El número debe ser un valor entero positivo mayor que 1. El valor predeterminado es 5.

Estado aleatorio de Python
El valor de random_state en Python se utiliza al realizar la evaluación de validación cruzada de modelos. Permite la réplica de resultados que implican números pseudoaleatorios. Debe ser un entero en el rango de 0 a 2^{32}-1. El valor predeterminado es 0.

Límite de tiempo (minutos)
El número de minutos permitidos para que se ejecuten los cálculos del modelo. Si especifica 0, el temporizador se desactiva. El valor predeterminado es 5.

Visualización

Esta sección especifica la cantidad de salida que se va a visualizar para la modalidad **Selección alfa mediante validación cruzada** .

El mejor
Sólo muestra los resultados básicos para el mejor modelo elegido. Se establece de forma predeterminada.

Comparar modelos
También muestra resultados básicos para todos los modelos evaluados.

Comparar modelos y pliegues
Por último, muestra resultados detallados completos para cada división o plegado para cada modelo evaluado.

Gráfico

Especifica gráficos de valores observados o residuales frente a valores pronosticados, con validación cruzada, especificación de gráficos de promedio de error cuadrático promedio (MSE) y/o promedio R^2 sobre pliegues de validación cruzada frente a valores alfa.

Promedio de error cuadrático promedio de validación cruzada (MSE) frente a alfa
La modalidad **Selección alfa mediante validación cruzada** muestra un gráfico de líneas de promedio de MSE sobre pliegues de validación cruzada frente a alfa. Para la modalidad de **Gráfico de rastreo** , se genera automáticamente un gráfico similar basándose en los datos de entrenamiento completos.

Promedio de R cuadrado de validación cruzada frente a alfa
La modalidad **Selección alfa mediante validación cruzada** muestra un gráfico de líneas de R^2 promedio sobre pliegues de validación cruzada frente a alfa. Para la modalidad de **Gráfico de**

rastreo , se genera automáticamente un gráfico similar basándose en los datos de entrenamiento completos.

Observado frente a Predicho
Muestra un diagrama de dispersión de los valores observados frente a los pronosticados para el modelo especificado o el mejor.

Residuos frente a pronosticados
También muestra un diagrama de dispersión de residuos frente a valores pronosticados para el modelo especificado o el mejor.

Guardar

Especifica variables para guardar en el conjunto de datos activo.

Valores pronosticados
Guarde los valores pronosticados del modelo especificado o mejor en el conjunto de datos activo. De lo contrario, puede especificar un **Nombre de variable personalizada**.

Residuos
Guarde los residuos de las predicciones de modelo especificadas o mejores en el conjunto de datos activo. De lo contrario, puede especificar un **Nombre de variable personalizada**.

Análisis de mediación

El análisis de mediación es un método estadístico que se utiliza para comprender el mecanismo por el cual una variable independiente influye en una variable dependiente a través de una o varias variables mediadoras. Este procedimiento te ayuda a comprobar hipótesis sobre efectos indirectos y a comprender las vías de influencia en tus datos.

El procedimiento de análisis de IBM SPSS Estadística mediación ofrece las siguientes funciones:

- Analizar los efectos directos e indirectos entre las variables predictoras, mediadoras y de resultado
- Desglosar los efectos totales en sus componentes para comprender las relaciones causales
- Gestionar diversos tipos de variables, incluidos datos continuos, binarios y de recuento
- Control de las variables de confusión mediante el ajuste de covariables
- Realizar análisis por subgrupos para examinar la heterogeneidad del efecto
- Visualizar las relaciones de mediación mediante diagramas de rutas y otros gráficos

Conceptos clave

- **Variable predictora (A)** : La variable independiente cuyo efecto se desea examinar (por ejemplo, la exposición al tratamiento, el nivel de intervención).
- **Variable mediadora (M)** : La variable intermedia que transmite el efecto de la variable predictora al resultado (por ejemplo, el estado psicológico o un marcador biológico).
- **Variable de resultado (Y)** : La variable dependiente que se ve afectada tanto por la variable predictora como por la mediadora (por ejemplo, un resultado de salud o un indicador de rendimiento).
- **Efecto directo** : el efecto de la variable predictora sobre el resultado que no se produce a través del mediador.
- **Efecto indirecto** : el efecto de la variable predictora sobre la variable dependiente que se produce a través del mediador.
- **Efecto total** : el efecto global de la variable predictora sobre la variable dependiente, que combina las vías directas e indirectas.

Aplicaciones
El análisis de mediación se utiliza ampliamente en diversas disciplinas:

- **Ciencias de la salud** : Comprender cómo las intervenciones influyen en los resultados de salud a través de mecanismos conductuales o biológicos

- **Psicología** : Análisis de cómo los tratamientos influyen en la salud mental a través de procesos cognitivos o emocionales
- **Ciencias sociales** : Análisis de cómo las políticas influyen en los resultados sociales a través de factores intermedios
- **Negocios** : Análisis de cómo las estrategias de marketing influyen en las ventas a través de las actitudes de los clientes
- **Educación** : Estudio de cómo los métodos de enseñanza influyen en los resultados del aprendizaje a través de la participación de los alumnos

Este procedimiento resulta especialmente útil cuando necesitas:

- Comprobar modelos teóricos sobre los mecanismos causales
- Identificar los objetivos de la intervención mediante el análisis de las vías de mediación
- Evaluar la importancia relativa de las diferentes vías causales
- Evaluar si los efectos varían entre los distintos subgrupos
- Control de las variables de confusión en los estudios observacionales

Tipos de análisis

El procedimiento admite dos métodos de análisis:

- **Análisis de mediación causal** : calcula efectos causales exhaustivos, incluidos el efecto directo controlado (CDE), el efecto directo natural (NDE), el efecto indirecto natural (NIE) y los componentes de la descomposición. Este enfoque tiene en cuenta las interacciones entre predictores y mediadores y ofrece un desglose detallado de los efectos.
- **Análisis de mediación tradicional** : calcula efectos de mediación más sencillos, sin términos de interacción ni descomposición detallada. Este enfoque es adecuado cuando no se esperan interacciones entre la variable predictora y la mediadora.

Combinaciones de variables admitidas

El procedimiento admite diferentes combinaciones de tipos de variables:

- Variable predictora continua, mediador continuo, resultado continuo
- Variable predictora continua, mediador binario, resultado continuo
- Variable predictora continua, mediador continuo, resultado binario
- Variable predictora continua, mediador binario, resultado binario
- Variable predictora continua, mediador continuo, variable de resultado discreta

Métodos estadísticos

El procedimiento utiliza métodos basados en la regresión para estimar los efectos de mediación:

- Regresión lineal para variables continuas
- Regresión logística para variables binarias
- Regresión de Poisson o regresión binomial negativa para variables de recuento

Los errores estándar y los intervalos de confianza se pueden calcular utilizando el método delta (predeterminado) o el re-muestreo bootstrap. El método delta permite realizar cálculos rápidos basados en la teoría asintótica, mientras que el bootstrapping ofrece una inferencia más robusta para distribuciones no normales o muestras pequeñas.

Estadísticos

El procedimiento ofrece resultados estadísticos completos, organizados en varias categorías:

Resúmenes del modelo de regresión

Modelo de regresión mediada:

- Coeficientes (β_0, β_1, β_2) con errores estándar y valores p
- Para mediadores continuos: $E(M|a,c) = \beta_0 + \beta_1 a + \beta_2 c$

- Para variables dicomáticas: modelo de regresión logística

Modelo de regresión de resultados:

- Coeficientes (θ_0, θ_1, θ_2, θ_3, θ_4) con errores estándar y valores p
- Para variables continuas: $E(Y|a, m, c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3(a \times m) + \theta_4 c$
- Para resultados binarios: modelo de regresión logística
- Para datos de recuento: regresión de Poisson o regresión binomial negativa

Medidas del efecto

Estadísticas comparativas de los niveles de exposición:

- **CDE (efecto directo controlado)** : efecto de la variable predictora sobre el resultado cuando el nivel del mediador se mantiene fijo
- **NDE (Efecto directo natural)** : Efecto directo que permite que el mediador varíe de forma natural
- **NIE (efecto indirecto natural)** : efecto mediado por un mediador
- Se presentan como diferencias en el caso de los resultados continuos, y como odds ratios (OR) o excRR en el caso de los resultados binarios

Resultados del análisis de descomposición:

- **TE (efecto total)** : efecto global de la variable predictiva sobre el resultado
- **PIE (efecto indirecto puro)** : componente del efecto indirecto
- **TDE (efecto directo total)** : componente del efecto directo
- **PAI (Interacción pura)** : Interacción atribuible al predictor
- **INTmed (Interacción mediada)** : Interacción a través de un mediador
- **INTref (Interacción de referencia)** : Efecto de interacción de referencia

Estadísticas de ratios (para resultados binarios):

- **PM (proporción mediada)** : proporción del efecto total que está mediada
- **Porcentaje (INT)** : Porcentaje debido a la interacción
- **Porcentaje (ELIM)** : Porcentaje eliminado

Métodos de inferencia

Método Delta (predeterminado):

- Errores estándar calculados a partir de las matrices de covarianza ($\Sigma\beta$, $\Sigma\theta$)
- Intervalos de confianza basados en la distribución normal
- Cálculo rápido, adecuado para la mayoría de los análisis

Método Bootstrap:

- Errores estándar e intervalos de confianza basados en el remuestreo
- Por defecto: 1000 muestras de bootstrap con un nivel de confianza del 95 %
- Intervalos percentiles o BCa (corregidos por sesgo y acelerados)
- Más robusto para distribuciones no normales

Opciones de la matriz de covarianza:

- **MODELO** : Utiliza la varianza estimada del modelo (resulta más adecuado cuando los datos se ajustan bien al modelo)
- **ROBUST** : Ajusta los valores atípicos y la heteroscedasticidad (recomendado para variables binarias)

Opciones de visualización

Diagrama de ruta:

- Representación gráfica de las vías de mediación
- Muestra los efectos directos e indirectos con los tamaños del efecto
- Visualización opcional de los efectos de las covariables

Diagrama de dispersión:

- Relaciones entre la variable predictora, la variable mediadora y la variable de resultado
- Líneas de tendencia (lineales o logísticas) con intervalos de confianza

Gráfico de barras:

- Recuentos o proporciones para variables categóricas
- Barras de error e intervalos de confianza

Gráfico de bosque:

- Resumen de las medidas del efecto con intervalos de confianza
- Muestra los coeficientes de ruta o las estimaciones de los efectos

Consideraciones sobre los datos

Tabla 1. Requisitos variables

Variable	Requisito
Variable de resultado	• Datos continuos, binarios o de recuento • Para resultados binarios: especifique la categoría de referencia (ÚLTIMO, PRIMERO o un VALOR concreto) • Para resultados de recuento: elige entre la distribución de Poisson o la distribución binomial negativa
Variable mediadora	• Datos continuos o binarios • Para variables continuas: especifique el nivel de análisis (MEDIA, MEDIANA o un VALOR concreto) • Para mediadores binarios: especificar la categoría de referencia
Variable predictora	• Datos continuos o categóricos • Especifique el nivel de referencia (BLEVEL) y el nivel de exposición (ELEVEL) para las comparaciones de efectos • Término opcional de interacción entre predictor y mediador (INTERACTION = TRUE/FALSE)

Especificaciones de modelos

El procedimiento utiliza el marco de los modelos lineales generalizados (GLM), lo que requiere especificar tanto la distribución como la función de enlace.

Para el modelo Mediator	Para el modelo de resultados
• **Una distribución normal con un enlace de identidad** implica una regresión lineal para los mediadores continuos	• **La distribución normal con enlace de identidad** implica una regresión lineal para los resultados continuos

Para el modelo Mediator	Para el modelo de resultados
• **La distribución binomial con enlace logit** implica una regresión logística para mediadores binarios	• **La distribución binomial con enlace logit** implica una regresión logística para resultados binarios • **La distribución de Poisson con enlace logarítmico** implica una regresión de Poisson para datos de recuento (supone que la varianza es igual a la media) • **Modelo binomial negativo con enlace logarítmico** para datos de recuento sobredispersos (la varianza supera la media) – **NEGBIN(1)** : Parámetro de dispersión fijo α = 1 – **NEGBIN(MLE)** : Estima el parámetro de dispersión a partir de los datos

El parámetro de dispersión α controla la sobredispersión: $Var(Y) = \mu + \alpha\mu^2$. Cuando α = 0, el modelo se reduce a la distribución de Poisson. Los valores de α más altos indican una mayor variabilidad con respecto a la media.

Tamaño de la muestra y convergencia

Tabla 2. Parámetros del algoritmo

Parámetro	Descripción	Rango
MÁXIMAS ITERACIONES	Número máximo de iteraciones	De 10 a 1000 Aumentar a 200-500 para modelos complejos
PCONVERGE	Tolerancia de convergencia	1E-8 a 1E-3 Utiliza 1E-8 para las publicaciones y 1E-4 para los análisis exploratorios
Singular	Umbral de detección de singularidades	valor numérico comprendido entre 1E-14 y 1E-6
MAXSTEPHALVING	Intentos de reducción a la mitad por pasos	Del 1 al 20
SANCIÓN POR SOBRESCRUCIAMIENTO	Regularización de crestas	1E-8 a 1E-3

Consideraciones de calidad de datos

Faltan datos

- Se excluyen los casos que presentan valores perdidos en las variables de análisis
- Informes resumidos sobre la tramitación de casos: casos válidos y excluidos

Ponderaciones

- Las ponderaciones de frecuencia se pueden aplicar mediante la variable «Scale Weight»
- El remuestreo bootstrap redondea las ponderaciones al entero más cercano

Covariables

- Incluir covariables para controlar los factores de confusión
- Tener en cuenta los factores que puedan influir tanto en el mediador como en el resultado

Análisis de subgrupos

- Utilice la cláusula BY para realizar análisis por separado según distintos factores (por ejemplo, sexo, grupo de edad)
- Útil para analizar la heterogeneidad de los efectos entre los distintos grupos

Supuestos

Supuestos del modelo

- No se observan factores de confusión no medidos en las relaciones entre predictor y mediador, entre predictor y resultado, y entre mediador y resultado
- Especificación correcta de las funciones de distribución y enlace
- Para la interpretación causal: orden temporal (predictor → mediador → resultado)
- No hay errores de medición en las variables

Supuestos estadísticos

- Tamaño de muestra adecuado para obtener estimaciones de parámetros estables
- Para el bootstrap: un número suficiente de muestras (se recomienda un mínimo de 1000)
- Para resultados binarios: número adecuado de casos por variable (se recomienda un mínimo de 10-15)
- En el caso de los modelos de interacción: potencia estadística suficiente para detectar efectos de interacción

Obtención del análisis de mediación

1. En los menús seleccione:

   Analizar > **Análisis de mediación**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

Tipo de modelo

En la pestaña «**Tipo de modelo** », defina la naturaleza estadística del análisis de mediación seleccionando el **«Tipo de mediador»** y **el «Tipo de resultado** ».

Seleccione el tipo de modelo de mediación que desea analizar. Las configuraciones de **«Tipo de mediador»** y **«Tipo de resultado»** determinan la distribución subyacente, la función de enlace y qué variables se permiten en las pestañas posteriores. Las opciones que selecciones en esta pestaña determinan qué opciones están activadas en las pestañas **«Variables dependientes»** y «**Criterios** ».

Los campos «Tipo de mediador» y «Tipo de resultado» son obligatorios.

Tipo de mediador

Selecciona cómo se modela la variable mediadora.

Continuo
Supone una distribución normal con una función de enlace identidad. Solo se admiten variables continuas para el mediador.

Utiliza esta opción cuando el mediador se mide en una escala continua.

Binario

Supone una distribución binomial con una función de enlace LOGIT. Solo se admiten variables binarias para el mediador.

Utiliza esta opción cuando el mediador tenga dos valores posibles.

Tipo de resultado

Selecciona cómo se modela la variable de respuesta. Las tablas de resultados se generan cuando la prevalencia de la variable de resultado es de al menos un 0.30 e.

Continuo

Supone una distribución normal con una función de enlace identidad.

Selecciona la opción para establecer el tipo de resultado continuo, lo que implica una distribución de datos NORMAL y una función de enlace IDENTIDAD. Si seleccionas esta opción, puedes configurar la **variable de resultado** en la pestaña **«Variables dependientes»** para que solo admita variables continuas.

Binario

Supone una distribución binomial. Puedes seleccionar la función de enlace **LOGIT** o **LOG**.

Selecciona la opción para establecer el tipo de resultado binario con la distribución binomial y la función de enlace logit. Si seleccionas esta opción, puedes configurar el campo **«Variable de resultado»** en la pestaña **«Variables dependientes»** y activar la opción «Binaria como resultado».

En la lista desplegable «Función de enlace», puedes seleccionar **«Logit»** (opción predeterminada) o «Función de enlace **logarítmica** para resultados binarios». Por defecto, la opción **«Binary»** establece la distribución binomial y la función de enlace LOGIT. En la pestaña **«Variables dependientes»**, la variable de resultado se puede especificar como **«Binaria** con **categoría de referencia»** para seleccionar entre valores binarios.

Recuentos

Selecciona esta opción para configurar la variable «counts» como **«Resultado»** en la pestaña «**Variables dependientes** ». Puedes añadir distribuciones **de Poisson** o **binomiales negativas** con una función de enlace logarítmica. Hay disponibles ajustes de parámetros adicionales para la opción «**Binomial negativa** ». Puedes ajustar el parámetro de dispersión α, que controla el grado de sobredispersión en el modelo. Un valor más alto de α indica una mayor variabilidad con respecto a la media.

- **Especificar valor** : Introduzca un número entero positivo fijo.
- **Valor estimado** : Selecciona esta opción si deseas estimar el parámetro de dispersión α mediante el método de máxima verosimilitud.

Variables dependientes

En la pestaña **«Variables dependientes»**, puede especificar la variable de resultado, que define el resultado del análisis de mediación.

Especifique la variable de respuesta (resultado), con opciones de personalización para variables continuas, binarias (incluida la categoría de referencia) y de recuento. Además, define la variable mediadora con opciones para variables continuas y binarias (incluida la categoría de referencia). Puedes configurar los niveles de control para el análisis de efectos y para incluir una variable de ponderación de escala.

También puede especificar la variable mediadora en el campo **«Variable mediadora»**, que define el componente mediador para el análisis de mediación. Todas las variables deben ser únicas tanto dentro de las variables dependientes e independientes como entre ellas y la variable «Ponderación de la escala».

Variable de resultado

Define la variable de resultado para el análisis de mediación. Proporcione variables continuas, binarias y de recuento. Esto depende del tipo de resultado que hayas especificado en la pestaña **«Tipo de modelo»**, siendo la variable continua la opción predeterminada. Si se opta por un resultado binario, puedes seleccionar variables binarias. Si se elige la distribución de Poisson o la distribución binomial negativa, también se pueden seleccionar variables de recuento.

Variable de resultado binario

Puedes seleccionar «binario» como resultado en la regresión de resultados. Si se selecciona un resultado binario, puede seleccionar el valor de referencia para la variable binaria seleccionada. Por defecto, una distribución binomial y una función de enlace logit/log.

Las tablas de resultados se generan cuando la prevalencia de la variable de resultado es de al menos un 0.30 e.

Categoría de referencia

Con esta opción, puede definir la categoría de referencia para el tipo de resultado binario que haya seleccionado en la pestaña «**Tipo de modelo** ». Seleccione cualquiera de las siguientes opciones:

- **Último (valor más alto)** : Establece el último valor o el valor más alto como referencia.
- **Primero (valor más bajo)** : Establece el primer valor o el valor más bajo como referencia.
- **Personalizado** : Selecciona esta opción para establecer una selección de referencia personalizada. Puede seleccionar los niveles de la variable binaria elegida en la variable de resultado o introducir un valor, siempre que dicho valor figure en la lista de variables.

Variable mediadora

Con esta opción, puedes definir la **variable mediadora** para el análisis de mediación. El tipo de datos debe ser continuo y binario.

Esta opción se basa en el **tipo de mediador** que has especificado en la pestaña **«Tipo de modelo»**; el valor predeterminado es una variable continua con una distribución normal y la función de enlace IDENTITY. Si se elige un mediador binario, las variables binarias pueden seleccionarse con una distribución binomial y una función de enlace LOGIT.

Establecer el nivel del mediador

Puedes seleccionar un nivel de mediador que se pueda aplicar al cálculo del efecto directo controlado. Seleccione cualquiera de las siguientes opciones:

- **Media** : Selecciona esta opción para establecer el nivel del mediador en la media de la variable mediadora seleccionada.
- **Mediana** : Selecciona esta opción para establecer el nivel del mediador en la mediana de la variable mediadora seleccionada.
- **Personalizado** : Selecciona esta opción para especificar un nivel personalizado. Esta opción solo se puede seleccionar para variables binarias. Introduce un valor numérico para el nivel del mediador. El valor debe estar presente en los valores observados del mediador binario. Si el mediador es binario, el usuario puede seleccionar el nivel binario (0 o 1).

Ponderación de escala

Selecciona la opción para asignar una variable de ponderación de escala. El campo solo admite variables con tipo de datos numérico.

Predictor

En la pestaña **«Predictor»** puedes especificar la variable predictora en el campo «Variable **predictora principal** », que define el componente **predictor** o **de exposición**, denotado por *A*, en el análisis de mediación.

Especifique la variable predictora principal junto con factores y covariables adicionales, con la posibilidad de establecer valores de referencia y de exposición para la variable predictora a fin de facilitar el análisis de efectos.

Los valores observados de la variable predictora en el conjunto de datos están representados por el vector a.

Predictor principal
Define la variable predictora principal o variable de exposición A para el análisis de mediación.

Solo se admiten variables continuas (se excluyen las variables categóricas para garantizar la claridad en la definición del efecto de la exposición).

Establecer nivel de línea base
Puede especificar los niveles de referencia mediante esta opción. Introduzca los valores en los campos: **«Establecer nivel de referencia»** y «**Establecer nuevo nivel de exposición** ». Ambos valores deben estar presentes en el conjunto de datos de la variable seleccionada como predictora.

- Establecer nivel de referencia: el nivel más bajo (o el primer valor) de la variable predictora en su orden natural.
- Establecer un nuevo nivel de exposición: el siguiente nivel superior (o segundo valor) de la variable predictora en el mismo orden.

Factores
Esta lista define los factores F que representan variables categóricas en el modelo. Cada factor se convierte en un conjunto de variables ficticias que se incluirán en la matriz C. Estas variables ficticias se incluyen en la regresión de resultados para tener en cuenta las diferencias entre grupos.

Se admiten variables categóricas (por ejemplo, grupos de tratamiento, sexo, etc.).

Los factores son opcionales, pero si el análisis incluye grupos categóricos o si se evalúan efectos de interacción, estas variables son obligatorias.

Covariables
Esta lista define las covariables *C* que controlan los posibles efectos de confusión en el modelo. Cada covariable se incluye en el modelo de regresión del resultado Y.

Si las covariables son categóricas, deben convertirse en variables ficticias antes de incluirlas en el modelo para evitar redundancias.

Cada covariable se incluye como efecto principal en el modelo de regresión del resultado Y.

Criterios

En la pestaña **«Criterios»**, puede definir opciones para seleccionar métodos de estimación de la matriz de covariables, calcular estimaciones de parámetros y especificar tipos de inferencia, el tipo de análisis y opciones para ajustar el algoritmo.

Matriz de covarianzas
Determina el método utilizado para estimar la matriz de covarianza de las estimaciones de los parámetros en el análisis de mediación.

- **Estimador basado en un modelo** : calcula la matriz de covarianza a partir de los supuestos del modelo estadístico especificado.
- **Estimador robusto** : proporciona una estimación robusta de la matriz de covarianza que se ajusta a las violaciones del modelo, como la heterocedasticidad o la falta de normalidad.

Tipo de inferencia
Esta opción sirve para especificar el método utilizado para la inferencia en el análisis de mediación.

- **Delta** : Este método utiliza la distribución asintótica del estimador para calcular los errores estándar y los intervalos de confianza.
- **Bootstrap** : Este método utiliza el remuestreo para crear una distribución del estimador con fines de inferencia, incluidos los intervalos de confianza.

  Opciones:

- **Número de muestras** : Para los intervalos percentiles y de intervalos de confianza (BCa) generados, se recomienda utilizar al menos 1000 muestras de bootstrap. Especifique un número entero positivo. El valor predeterminado es 500.
- **Establecer la semilla para Mersenne Twister** : Establecer una semilla te permite repetir los análisis. El uso de este control es parecido a establecer el tornado de Mersenne como generador activo y especificar un punto de inicio fijo en el cuadro de diálogo Generadores de números aleatorios, con la importante diferencia de que la definición de la semilla de este cuadro de diálogo mantendrá el estado actual del generador de números aleatorios y restaurará dicho estado cuando haya terminado el análisis.
- **Intervalos de confianza** : especifique un nivel de confianza superior a 50 e inferior a 100. Los intervalos de percentiles sólo utilizan los valores de simulación de muestreo ordenados correspondientes a los percentiles de intervalo de confianza deseados. Por ejemplo, un intervalo de confianza de percentil del 95% utiliza los percentiles 2,5 y 97,5 de los valores de simulación de muestreo como los límites inferior y superior del intervalo (interpolando los valores de simulación de muestreo si es necesario). Los intervalos de sesgo corregidos y acelerados (BCa) son intervalos ajustados que son más precisos, pero necesitan más tiempo de cálculo.

Tipo de análisis

- **Causal** : Calcula los efectos completos de mediación causal, incluyendo el CDE, el NDE, el NIE, los efectos totales, las razones de probabilidades, los riesgos relativos excesivos y las descomposiciones de interacción.
- **Tradicional** : calcula efectos de mediación más sencillos sin términos de interacción ni descomposición causal completa.

Opciones del algoritmo

Las opciones están habilitadas cuando el «Tipo de resultado» es «**Binario** » o **«Recuento»** en la pestaña «**Tipo de** modelo».

- **Número máximo de iteraciones** : establece el número máximo de iteraciones del algoritmo IRWLS antes de que este se detenga. Puedes introducir valores enteros comprendidos entre 10 y and1000. El valor predeterminado es 100.
- **Tolerancia de convergencia** : establece el umbral de convergencia de los parámetros (variación absoluta de la norma del vector de coeficientes).

  Umbral de convergencia de los parámetros. El valor predeterminado es 1E-6. El algoritmo se detiene cuando la variación absoluta de los coeficientes es inferior a este valor. Utiliza 1E-8 para obtener una precisión apta para publicaciones y 1E-4 para realizar análisis exploratorios más rápidos.

  Introduzca un valor numérico comprendido entre 1E-8 y 1E-3.

  El campo admite formatos de notación científica: 1E-6, 1e-6, 0.000001
- **Tolerancia a la singularidad** : establece el umbral para detectar matrices Hessianas singulares (umbral del determinante). Umbral para detectar matrices hessianas singulares. El valor predeterminado es 1E-12. El algoritmo detecta una situación cercana a la singularidad cuando el determinante de la matriz hessiana cae por debajo de este valor. Utiliza 1E-10 para comparar los resultados de R y Stata. Consulte los siguientes requisitos previos para la tolerancia a singularidades:
  - Los valores más bajos son más conservadores (menos detecciones erróneas de singularidades).
  - Introduzca un valor numérico comprendido entre 1E-14 y 1E-6
  - El campo admite formatos de notación científica: 1E-12, 1e-12, 0.000000000001
- **Reducción máxima a la mitad** :

  Este campo indica el número máximo de intentos de reducción a la mitad del paso por iteración. El valor predeterminado es 5. La reducción a la mitad del paso garantiza que la verosimilitud logarítmica mejore en cada iteración. Aumenta el valor a 10-20 si se producen errores de reducción a la mitad del paso. El valor debe ser un número entero comprendido entre 1 y 20.

- **Penalización de Ridge** : Establece la intensidad de la regularización de Ridge para matrices Hessianas singulares.

  El valor predeterminado es 1E-6. Cuando la matriz Hessiana es casi singular (con un determinante inferior al umbral de tolerancia a la singularidad), se aplica esta penalización para estabilizar la inversión de la matriz. Utiliza valores más altos (1E-4, 1E-3) para las variables predictoras con alta colinealidad.

  Introduzca un valor numérico comprendido entre 1E-8 y 1E-3. El campo también admite formatos de notación científica: 1E-6, 1e-6, 0.000001

Imprimir

En la pestaña **«Imprimir»**, puedes configurar opciones para personalizar la salida e incluir solo los componentes relevantes que selecciones.

Seleccione cualquiera de las siguientes opciones para generar la sección de resultados correspondiente como parte de los resultados del análisis.

Resumen del modelo de regresión de mediación
: Esta opción está seleccionada de forma predeterminada como opción integrada y no se puede modificar. En la salida se genera una tabla que resume los resultados del modelo de regresión de mediación. Incluye los coeficientes estimados (β0, β1, β2), sus errores estándar y los niveles de significación estadística. Estos resultados ayudan a comprender la relación entre la variable predictora (*A*), la variable mediadora (*M*) y las covariables (*C*). Esta opción se selecciona de forma predeterminada.

Resumen del modelo de regresión de resultados
: Esta opción está seleccionada de forma predeterminada como opción integrada y no se puede modificar. En la salida se genera una tabla que resume los resultados del modelo de regresión de resultados. Proporciona los coeficientes estimados (θ0, θ1, θ2, θ3, θ4), sus errores estándar y los niveles de significación. Estos resultados evalúan los efectos de la variable predictora, la variable mediadora y las covariables sobre la variable dependiente (*Y*). Esta opción se selecciona de forma predeterminada.

Estadísticas de comparación de los niveles de exposición
: Esta opción está seleccionada de forma predeterminada como opción integrada y no se puede modificar. En el resultado se genera una sección que muestra las diferencias en los resultados previstos según los distintos niveles de exposición de la variable predictora. Las estadísticas incluyen estimaciones como el efecto directo controlado (*CDE*), el efecto directo natural (*NDE*) y el efecto indirecto natural (*NIE*), lo que permite comprender cómo los cambios en la exposición afectan al mediador y al resultado.

Resultados del análisis de descomposición
: Seleccione esta opción para generar una tabla que muestre el desglose del efecto total (*ET*) en componentes directos e indirectos, incluyendo *el CDE*, *el NDE* y *el NIE*. También presenta componentes adicionales, como la proporción mediada (*PM*) y los efectos de interacción.

Análisis de efectos de interacción
: Seleccione esta opción para generar una sección que evalúe efectos de interacción, como *INTmed* e *INTref*, que miden las contribuciones de las interacciones mediador-resultado y predictor-mediador. Estos resultados ayudan a comprender cómo influyen las interacciones en el resultado.

Estadísticos de la razón
: Seleccione esta opción para generar estadísticas de ratios como *r(CDE)*, *r(NDE)*, *r(NIE)* y *r(INTmed)*. Estos valores representan la proporción del efecto total (*ET*) atribuible a cada componente, expresada en porcentajes.

Matriz de covarianza para estimaciones de parámetros
: Esta opción se selecciona de forma predeterminada. Se puede generar una matriz que muestre las covarianzas entre las estimaciones de los parámetros, como *β* (de la regresión de mediación) y *ϑ* (de la regresión de resultados). Es fundamental para calcular los errores estándar de las estadísticas derivadas utilizando el método delta. Esta opción se selecciona de forma predeterminada.

Resumen de procesamiento de casos
Esta opción se selecciona de forma predeterminada. Puede generar un resumen que ofrezca una visión general de los datos utilizados en el análisis, incluyendo el número de casos, los casos excluidos y las ponderaciones aplicadas. Ayuda a evaluar la exhaustividad y la calidad del conjunto de datos. Esta opción se selecciona de forma predeterminada.

Estadísticos descriptivos
Seleccione esta opción para generar una sección que presente estadísticas descriptivas de todas las variables incluidas en el análisis, como medias, desviaciones estándar, valores mínimos y valores máximos. Ofrece una visión general de la distribución y la variabilidad de los datos.

Gráficos

En la pestaña **«Gráficos»**, puedes configurar opciones para generar visualizaciones que ayuden a explicar las relaciones entre las variables mediadoras, predictoras y de resultado.

Hay opciones disponibles para definir cuatro tipos de visualizaciones: diagramas de ruta, diagramas de dispersión, gráficos de barras y diagramas de bosque. Cada tipo de gráfico ofrece una perspectiva diferente del modelo de mediación y se puede personalizar para adaptarlo a tus necesidades de análisis.

Puedes personalizar cada gráfico añadiendo tamaños del efecto, intervalos de confianza o desgloses por subgrupos.

Diagrama de ruta
Visualiza las relaciones entre las variables mediadoras, predictoras y de resultado. Los diagramas de ruta son útiles para presentar el modelo de mediación global, incluidos los efectos directos e indirectos.

- **Ocultar efectos de covariables** : oculta todos los bordes asociados a las covariables. Las rutas «predictor-mediador», «mediador-resultado» y «predictor-resultado» siguen siendo visibles.
- **Color de los bordes** : cuando se selecciona esta opción, los bordes se codifican por colores para indicar el tipo de efecto (por ejemplo, efectos directos, indirectos y de covariables). El color verde indica efectos más intensos, mientras que el gris indica efectos más débiles. Si no se selecciona esta opción, el color predeterminado del borde será uniforme.

Diagrama de dispersión
Los diagramas de dispersión muestran las relaciones entre variables y ayudan a identificar correlaciones y patrones de distribución.

En el análisis de mediación, los diagramas de dispersión ayudan a visualizar las relaciones entre la variable predictora (X), el mediador (M) y la variable dependiente (Y). Estas relaciones se corresponden con las siguientes vías de mediación:

- Ruta a: Efecto de la variable predictora (X) sobre la variable mediadora (M)
- Ruta b: Efecto del mediador (M) sobre el resultado (Y)
- Ruta c′: Efecto directo de la variable predictora (X) sobre la variable dependiente (Y)

Un diagrama de dispersión de tres variables muestra cómo la variable predictora influye tanto en la variable mediadora como en el resultado, lo que ilustra el papel de la variable mediadora en la relación entre la variable predictora y el resultado. La visualización suele incluir:

- Puntos de datos que representan la relación entre la variable predictora y la mediadora (ruta a)
- Datos que muestran la relación entre el mediador y el resultado (ruta b)
- La relación global entre la variable predictiva y la variable de resultado, incluyendo tanto los efectos directos como los indirectos (ruta c′)

Con los diagramas de dispersión, puedes evaluar visualmente la intensidad, la dirección y la importancia de estas relaciones. Los ejes suelen representar las variables mediadoras y de resultado, mientras que la variable predictora se muestra mediante degradados de color o marcadores.

Gráfico de barras

Muestran proporciones o recuentos para las variables categóricas y resultan útiles para visualizar los niveles de los mediadores y los resultados. Esta opción solo está disponible cuando la variante de mediación seleccionada requiere un gráfico de barras (por ejemplo, mediador binario, resultado binario o mediador continuo con resultado binario).

La opción está habilitada para variantes con mediador binario, resultado binario o mediador continuo con resultado binario.

Esta opción está desactivada para las variantes de mediador continuo y de resultado continuo, en las que se utilizan diagramas de dispersión en su lugar.

- **Barras de error** : añade barras de error a los gráficos de barras. En los gráficos de barras agrupadas, las barras de error representan la incertidumbre en las proporciones (Y = 1 o Y = 0) para cada nivel del mediador.
- **Intervalos de confianza** : muestra los intervalos de confianza de las barras de error. Esto viene determinado por el valor que introduzcas en el campo **«Nivel de intervalo de confianza %»** de la pestaña «**Criterios** ». El 95 % es el valor predeterminado para el nivel del intervalo de confianza.

Gráfico de bosque

Muestra los coeficientes de ruta y sus intervalos de confianza, que te ayudan a evaluar la magnitud y la incertidumbre de los efectos de mediación.

Los gráficos de bosque muestran los tamaños del efecto de las rutas de mediación (por ejemplo, a, b y c′) junto con sus intervalos de confianza correspondientes.

Opciones:

- **Coeficientes de ruta** : muestra los coeficientes de ruta (a, b y c′) para cada ruta de mediación.
- **Intervalos de confianza** : muestra los intervalos de confianza para cada coeficiente de ruta. Esto viene determinado por el valor que introduzcas en el campo **«Nivel de intervalo de confianza %»** de la pestaña «**Criterios** ». El 95 % es el valor predeterminado para el nivel del intervalo de confianza.

Filtros de series temporales

En el análisis de series temporales, es esencial eliminar los valores que se salen de un rango decidido y provocan una fluctuación inusual en la serie temporal. Un buen filtro debe ser capaz de eliminar las raíces unitarias y los componentes cíclicos, o el filtro debe ser capaz de aislar las fluctuaciones de los datos a una determinada frecuencia.

El filtrado de series temporales es una técnica vital en econometría y análisis de series temporales, a menudo empleada para descomponer una serie temporal en componentes de tendencia y cíclicos. Esto es especialmente importante en macroeconomía y finanzas, donde es esencial identificar las fluctuaciones del ciclo económico o las tendencias económicas a largo plazo.

Los filtros de series temporales más utilizados son el filtro Hodrick-Prescott (HP), el filtro Baxter-King (BK) y el filtro Christiano-Fitzgerald (CF).

Obtención de filtros de series temporales

Los Filtros de Series Temporales están disponibles como procedimiento de extensión con el plug-in Python. Por lo tanto, instale el módulo de ampliación para ejecutar el procedimiento.

complete los siguientes pasos para ejecutar el procedimiento para Filtros de Series Temporales.

1. Vaya a **DatosDefinir** > **fecha y hora** para generar las variables de fecha.
2. Después de definir la fecha y la hora, haga clic en **Analice** > **Previsión** > **Filtros de series temporales (TSF)**.
3. Seleccione el filtro de series temporales que desea aplicar al conjunto de datos.

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

4. Seleccione las variables y muévalas a la lista de variables para el filtro de series temporales apropiado.
5. Establezca el valor de **Lambda** para el filtro Hodrick-Prescott (HP) o establezca los valores de **Bajo**, **Alto** y **K** para el filtro Baxter-King (BK) y el filtro Christiano-Fitzgerald (CF).

   Opcionalmente, puede seleccionar **Deriva** para el filtro Christiano-Fitzgerald (CF).
6. Pulse **Ejecutar**.

Filtros de series temporales: Opciones

El cuadro de diálogo Filtros de series temporales contiene las siguientes opciones.

Importante: Antes de configurar las opciones, vaya a **DatosDefinir** > **fecha y hora** para generar las variables de fecha.

Hodrick-Prescott (HP)

El filtro **de Hodrick-Prescott (HP)** es una herramienta utilizada en macroeconomía para ajustar una tendencia suave a series temporales.

El filtro de Hodrick-Prescott separa una serie temporal $_{yt}$ en un componente tendencial $_{\tau t}$ y un componente cíclico $_{ct}$.

$y_t = \tau_t + c_t + \square_t$donde $\square_{es}$ el componente de error (o ruido).

Los componentes se determinan minimizando la función de pérdida cuadrática y el filtro HP viene dado explícitamente por la fórmula,

$HP = [\ ^{\lambda L2-}\ 4\lambda L + (1 + 6\lambda)-\ ^{4\lambda L-1} + \ ^{\lambda L-2}\ ]^{-1}$

donde λ denota el parámetro de suavizado (valor positivo) y L denota el operador de retardo, que se deriva de la condición de primer orden para el problema de minimización.

Elija una variable para el filtro HP de la lista de variables y establezca un valor para **Lambda**.

El parámetro de suavizado *λ* determina el equilibrio entre ajuste y suavidad. Cuanto mayor sea el valor de *λ*, mayor será la penalización. Cuando se analizan datos trimestrales, se recomienda el valor por defecto de 1600 para *λ*. Además, *λ* debe ser $^{1600/44}$ = 6.25 para los datos anuales, y 1600 $*\ ^{34}$ = 129.600 para los datos mensuales. Cuando *λ*= 0, significa que la salida de la serie del filtro HP coincide exactamente con la serie de entrada. Por el contrario, valores más altos de *λ* producen series de salida más suaves.

Baxter-King (BK)

Baxter-King (BK) es un filtro pasa banda utilizado habitualmente en el análisis de series temporales para aislar los componentes cíclicos de los datos, a menudo en estudios macroeconómicos. Su objetivo es eliminar el ruido de alta frecuencia (a corto plazo) y los componentes de baja frecuencia (tendencia a largo plazo), centrándose en los ciclos de interés a medio plazo, normalmente entre 1.5 y 8 años. El filtro BK asume la estacionariedad de la serie temporal subyacente. En la lista de variables, elija 1 o 2 variables para el filtro BK.

Establezca los valores adecuados en **Bajo** (periodo mínimo), **Alto** (periodo máximo) y **K** (longitud de Lead-lag).

Christiano-Fitzgerald (CF)

El filtro **Christiano-Fitzgerald (CF)** es otro filtro de paso de banda utilizado para separar una serie temporal en componentes de tendencia y cíclicos. Este filtro es similar al filtro Baxter King (BK), que también funciona sobre la base de un filtro pasa banda ideal donde el filtro BK asume una media móvil simétrica para aproximarse al filtro pasa banda, pero el filtro CF asume que la serie temporal sigue la aleatoriedad sin la deriva. El filtrado CF se implementa estimando un filtro óptimo bajo el supuesto de que los datos son generados por un paseo aleatorio o un proceso de tipo ARIMA.

En la lista de variables, elija 1 o 2 variables para el filtro CF. Establezca los valores adecuados en **Bajo**, **Alto** y **K**. Seleccione **Deriva** si es necesario para el filtro CF.

Lambda, **Low** y **High** pueden ser cualquier valor flotante. El valor de **K** debe ser un número entero. Consulte los siguientes valores:

Bajo	1.5 (anual) y 6 (trimestral)
Alto	8 (anual) y 32 (trimestral)
K	3 (anual) y 12 (trimestral)

Estimación curvilínea

El procedimiento Estimación Curvilínea genera estadísticos de estimación curvilínea por regresión y gráficos relacionados para 11 modelos diferentes de estimación curvilínea por regresión. Se produce un modelo diferente para cada variable dependiente. También se pueden guardar valores predichos, residuos e intervalos de predicción como nuevas variables.

Ejemplo. Un proveedor de servicios de Internet realiza un seguimiento del porcentaje de tráfico de correo electrónico infectado de virus en la red a lo largo del tiempo. Un diagrama de dispersión revela que la relación es no lineal. Se puede ajustar un modelo lineal a los datos y comprobar la validez de los supuestos y la bondad de ajuste del modelo.

Estadísticas. Para cada modelo: coeficientes de regresión, *R* múltiple, *R* cuadrado, *R* cuadrado corregida, error estándar de la estimación, tabla de análisis de varianza, valores pronosticados, residuos e intervalos de predicción. Modelos: lineal, logarítmico, inverso, cuadrático, cúbico, de potencia, compuesto, curva-S, logístico, de crecimiento y exponencial.

Estimación curvilínea: Consideraciones sobre los datos

Datos. Las variables dependiente e independientes deben ser cuantitativas. Si selecciona **Tiempo** del conjunto de datos activo como variable independiente (en lugar de una variable), el procedimiento Estimación curvilínea generará una variable de tiempo en la que la distancia temporal entre los casos es uniforme. Si se selecciona **Tiempo**, la variable dependiente debe ser una medida de serie temporal. El análisis de series temporales requiere una estructura particular para los archivos de datos, de manera que cada caso (cada fila) represente un conjunto de observaciones en un momento determinado del tiempo y que la distancia temporal entre los casos sea uniforme.

Supuestos. Represente los datos gráficamente para determinar cómo se relacionan las variables dependientes e independiente (linealmente, exponencialmente, etc.). Los residuos de un buen modelo deben distribuirse de forma aleatoria y normal. Si se utiliza un modelo lineal, se deben cumplir los siguientes supuestos: para cada valor de la variable independiente, la distribución de la variable dependiente debe ser normal. La varianza de distribución de la variable dependiente debe ser constante para todos los valores de la variable independiente. La relación entre la variable dependiente y la variable independiente debe ser lineal y todas las observaciones deben ser independientes.

Para obtener una estimación curvilínea

1. En los menús seleccione:

   Analizar > **Regresión** > **Estimación curvilínea...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables dependientes. Se produce un modelo diferente para cada variable dependiente.
3. Seleccione una variable independiente (seleccione una variable del conjunto de datos activo o **Tiempo**).
4. Si lo desea:

- Seleccionar una variable para etiquetar los casos en los diagramas de dispersión. Para cada punto en el diagrama de dispersión, se puede utilizar la herramienta de Identificación de puntos para mostrar el valor de la variable utilizada en Etiquetas de caso.

- Pulsar en **Guardar** para guardar los valores pronosticados, los residuos y los intervalos de predicción como nuevas variables.

También se encuentran disponibles las siguientes opciones:

- **Incluir la constante en la ecuación.** Estima un término constante en la ecuación de regresión. La constante se incluye de forma predeterminada.
- **Representar los modelos.** Representa los valores de la variable dependiente y cada modelo seleccionado frente a la variable independiente. Se genera un gráfico distinto para cada variable dependiente.
- **Ver tabla de ANOVA.** Muestra una tabla de análisis de varianza de resumen para cada modelo seleccionado.

Modelos del procedimiento Estimación curvilínea

Se puede seleccionar uno o más modelos de estimación curvilínea por regresión. Para determinar qué modelo utilizar, represente los datos. Si las variables parecen estar relacionadas linealmente, utilice un modelo de regresión lineal simple. Cuando las variables no estén relacionadas linealmente, intente transformar los datos. Cuando la transformación no resulte útil, puede necesitar un modelo más complicado. Inspeccione un diagrama de dispersión de los datos; si el diagrama se parece a una función matemática reconocible, ajuste los datos a ese tipo de modelo. Por ejemplo, si los datos se parecen a una función exponencial, utilice un modelo exponencial.

Lineal. Modelo cuya ecuación es Y = b0 + (b1 * t). Los valores de la serie se modelan como una función lineal del tiempo.

Logarítmico. Modelo cuya ecuación es Y = b0 + (b1 * ln(t)).

Inverso. Modelo cuya ecuación es Y = b0 + (b1 / t).

Cuadrático. Modelo cuya ecuación es Y = b0 + (b1 * t) + (b2 * t**2). El modelo cuadrático puede utilizarse para modelar una serie que "despega" o una serie que se amortigua.

Cúbico. Modelo definido por la ecuación Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3).

Potencia. Modelo cuya ecuación es Y = b0 * (t**b1) ó ln(Y) = ln(b0) + (b1 * ln(t)).

Compuesto. Modelo cuya ecuación es Y = b0 * (b1**t) o ln(Y) = ln(b0) + (ln(b1) * t).

Curva-S. Modelo cuya ecuación es Y = e**(b0 + (b1/t)) o ln(Y) = b0 + (b1/t).

Logística. Modelo cuya ecuación es Y = 1/(1/u + (b0 * (b1**t))) o ln (1/y-1/u) = ln (b0) + (ln (b1) * t) donde u es el valor de límite superior. Después de seleccionar Logístico, especifique un valor para el límite superior que se utilizará en la ecuación de regresión. El valor debe ser un número positivo mayor que el valor máximo de la variable dependiente.

Crecimiento. Modelo cuya ecuación es Y = e**(b0 + (b1 * t)) ó ln(Y) = b0 + (b1 * t).

Exponencial. Modelo cuya ecuación es Y = b0 * (e**(b1 * t)) ó ln(Y) = ln(b0) + (b1 * t).

Estimación curvilínea: Guardar

Guardar variables. Para cada modelo seleccionado se pueden guardar los valores pronosticados, los residuos (el valor observado de la variable dependiente menos el valor predicho por el modelo) y los intervalos de predicción (sus límites superior e inferior). En la ventana de resultados, se muestran en una tabla los nombres de las nuevas variables y las etiquetas descriptivas.

Pronosticar casos. En el conjunto de datos activo, si se selecciona **Tiempo** como variable independiente en lugar de una variable, se puede especificar un período de previsión que vaya más allá del final de la serie temporal. Puede elegir una de las siguientes alternativas:

- **Desde el período de estimación hasta el último caso.** Pronostica los valores para todos los casos del archivo, basándose en los casos del período de estimación. El período de estimación, que se muestra en la parte inferior del cuadro de diálogo, se define con el subcuadro de diálogo Rango de la opción

Seleccionar casos en el menú Datos. Si no se ha definido un período de estimación, se utilizan todos los casos para pronosticar los valores.

- **Predecir hasta.** Predice los valores hasta la fecha especificada, hora o número de observación, basándose en los casos del período de estimación. Esta característica se puede utilizar para prever valores más allá del último caso de la serie temporal. Las variables definidas actualmente determinan los cuadros de texto disponibles para especificar el final del período de predicción. Si no existen variables de fecha definidas, se puede especificar el número de la observación (caso) final.

Utilice la opción de Definir fechas en el menú Datos para crear las variables de fecha.

Regresión por mínimos cuadrados parciales

El procedimiento Regresión de mínimos cuadrados parciales estima los modelos de regresión de los mínimos cuadrados parciales (PLS, también denominados "proyección a la estructura latente"). La PLS es una técnica de predicción alternativa a la regresión de mínimos cuadrados ordinarios (OLS), a la correlación canónica o al modelado de ecuaciones estructurales, y resulta particularmente útil cuando las variables predictoras están muy correlacionadas o cuando el número de predictores es superior al número de casos.

La PLS combina las características del análisis de componentes principales y la regresión múltiple. En primer lugar, extrae un conjunto de factores latentes que explica en la mayor medida posible la covarianza entre las variables dependientes e independientes. A continuación, un paso de regresión pronostica los valores de las variables dependientes mediante la descomposición de las variables independientes.

Tablas
La proporción de la varianza explicada (por factor latente), las ponderaciones y las cargas de los factores latentes, la importancia de la variable independiente en proyección (VIP) y las estimaciones de los parámetros de la regresión (por variable dependiente) se generan de forma predeterminada.

Gráficos
La variable independiente en proyección (VIP), las puntuaciones factoriales, las ponderaciones factoriales de los tres primeros factores latentes y la distancia al modelo se generan desde la pestaña Options.

Consideraciones sobre los datos

Nivel de medición
Las variables (predictoras) dependientes e independientes pueden ser de escala, nominales u ordinales. El procedimiento supone que se ha asignado el nivel de medición adecuado a todas las variables, aunque puede cambiar temporalmente el nivel de medición de una variable pulsando el botón derecho la variable en la lista de variables de origen y seleccionando un nivel de medición en el menú emergente. El procedimiento trata por igual las variables categóricas (nominales u ordinales).

Codificación de variables categóricas
El procedimiento recodifica temporalmente las variables dependientes categóricas utilizando la codificación "una de *c*" durante el procedimiento. Si hay categorías *c* de una variable, la variable se almacena como vectores *c* , con la primera categoría indicada (1,0, ..., 0), la siguiente categoría (0,1,0, ..., 0), ..., y la categoría final (0,0, ..., 0, 1). Las variables dependientes categóricas se representan mediante una variable auxiliar o dummy; es decir, omitiendo simplemente el indicador correspondiente a la categoría de referencia.

Ponderaciones de frecuencia
Los valores de ponderación se redondean al número entero más cercano antes de utilizarlos. Los casos con ponderaciones perdidas o ponderaciones inferiores a 0,5 no se emplearán en los análisis.

Valores perdidos
Los valores perdidos del usuario y del sistema se consideran no válidos.

Cambio de escala
Todas las variables del modelo se centran y tipifican, incluidas las variables indicador que representan variables categóricas.

Obtención de regresión por mínimos cuadrados parciales

En los menús seleccione:

Analizar > **Regresión** > **Mínimos cuadrados parciales...**

1. Seleccione al menos una variable dependiente.
2. Seleccione al menos una variable independiente.

Si lo desea, puede:

- Especificar una categoría de referencia para las variables dependientes categóricas (nominales u ordinales).
- Especificar la variable que se utilizará como identificador exclusivo para los resultados por casos y los conjuntos de datos guardados.
- Especificar un límite máximo para el número de factores latentes que se extraerán.

Modelo

Especificar efectos del modelo. El modelo de efectos principales contiene todos los efectos principales de los factores y de las covariables. Seleccione **Personalizado** para especificar las interacciones. Indique todos los términos que desee incluir en el modelo.

Factores y Covariables. Muestra una lista de los factores y las covariables.

Modelo. El modelo depende de la naturaleza de los datos. Después de seleccionar **Personalizado**, puede elegir los efectos principales y las interacciones que sean de interés para el análisis.

Construir términos

Para las covariables y los factores seleccionados:

Interacción. Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el valor predeterminado.

Efectos principales. Crea un término de efectos principales para cada variable seleccionada.

Todas de 2. Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todas de 3. Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todas de 4. Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todas de 5. Crea todas las interacciones quíntuples posibles de las variables seleccionadas.

Opciones

La pestaña Opciones permite al usuario guardar y representar las estimaciones de los modelos para los determinados casos, factores latentes y predictores.

Para cada tipo de datos, especifique el nombre del conjunto de datos. Los nombres de los conjuntos de datos deben ser exclusivos. Si introduce el nombre de un conjunto de datos ya existente, se reemplazarán los contenidos. En otro caso, se creará un nuevo conjunto de datos.

- **Guardar estimaciones para casos individuales.** Guarda las siguientes estimaciones de modelos por casos: valores pronosticados, residuos, distancia respecto al modelo del factor latente y puntuaciones de los factores latentes. También representa las puntuaciones de los factores latentes.
- **Guardar estimaciones para factores latentes.** Guarda las cargas y las ponderaciones de los factores latentes. También representa las ponderaciones de factores latentes.
- **Guardar estimaciones para variables independientes.** Guarda las estimaciones de los parámetros de regresión y la importancia de la variable en la proyección (VIP). También representa la VIP por factor latente.

Análisis de vecinos más cercanos

Análisis de vecinos más próximos es un método para clasificar casos basándose en su parecido a otros casos. En el aprendizaje automático, se desarrolló como una forma de reconocer patrones de datos sin la necesidad de una coincidencia exacta con patrones o casos almacenados. Los casos parecidos están próximos y los que no lo son están alejados entre sí. Por lo tanto, la distancia entre dos casos es una medida de disimilaridad.

Los casos próximos entre sí se denominan “vecinos”. Cuando se presenta un nuevo caso (reserva), se calcula su distancia con respecto a los casos del modelo. Las clasificaciones de los casos más parecidos (los vecinos más próximos) se cuadran y el nuevo caso se incluye en la categoría que contiene el mayor número de vecinos más próximos.

Puede especificar el número de vecinos más cercanos a examinar; este valor se denomina *k*.

El método Análisis de vecinos más próximos también puede utilizarse para calcular valores para un destino continuo. En esta situación, la media o el valor objetivo medio de los vecinos más próximos se utiliza para obtener el valor predicho del nuevo caso.

Análisis de vecino más próximo: Consideraciones sobre los datos

Objetivo y características. El objetivo y las características pueden ser:

- *Nominal*. Una variable se puede tratar como nominal cuando sus valores representan categorías sin ninguna clasificación intrínseca (por ejemplo, el departamento de la empresa en la que trabaja un empleado). Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa.
- *Ordinal*. Una variable se puede tratar como ordinal cuando sus valores representan categorías con alguna clasificación intrínseca (por ejemplo, niveles de satisfacción del servicio de muy insatisfecho a muy satisfecho). Entre los ejemplos de variables ordinales se incluyen escalas de actitud que representan el grado de satisfacción o confianza y las puntuaciones de evaluación de las preferencias.
- *Escalar*. Una variable se puede tratar como escala (continua) cuando sus valores representan categorías ordenadas con una métrica significativa, de modo que las comparaciones de distancia entre valores son adecuadas. Son ejemplos de variables de escala: la edad en años y los ingresos en dólares.

  El análisis de vecinos más próximos trata por igual las variables nominales u ordinales. El procedimiento supone que se ha asignado el nivel de medición adecuado a cada variable. No obstante, puede cambiar temporalmente el nivel de medición para una variable pulsando con el botón derecho en la variable en la lista de variables de origen y seleccionar un nivel de medición en el menú emergente.

Un icono situado junto a cada variable de la lista de variables identifica el nivel de medición y el tipo de datos.

Tabla 3. Iconos de nivel de medición

	Numérico	**Serie**	**Fecha**	**Hora**
Escala (Continuo)		n/a		
Ordinal				
Nominal				

Codificación de la variable categórica. El procedimiento recodifica temporalmente predictores categóricos y variables dependientes utilizando la codificación "una de *c*" para todo el procedimiento. Si hay *c* categorías de una variable, entonces la variable se almacena como c vectores, con la primera categoría denotada (1,0,...,0), la siguiente categoría (0,1,0,...,0),..., y la categoría final (0,0,...,0,1).

Este esquema de codificación aumenta la dimensionalidad del espacio de características. En concreto, el número total de dimensiones es el número de predictores de escala más el número de categorías en todos los predictores categóricos. Como resultado, este esquema de codificación puede conllevar un entrenamiento más lento. Si el entrenamiento de vecinos más próximos avanza muy lentamente, pruebe a reducir el número de categorías en los predictores categóricos combinando categorías similares o eliminando los casos que tengan categorías extremadamente raras antes de ejecutar el procedimiento.

Toda codificación "una de *c*" se basa en los datos de entrenamiento, incluso si se define una muestra reservada (consulte “Particiones ” en la página 211). De este modo, si las muestras reservadas contienen casos con categorías de predictores que no están presentes en los datos de entrenamiento, esos casos no se puntúan. Si las muestras reservadas contienen casos con categorías de variables dependientes que no están presentes en los datos de entrenamiento, esos casos se puntúan.

Cambio de escala. Fe forma predeterminada, las características de escala se normalizan. Todo cambio de escala se realiza basándose en los datos de entrenamiento, incluso si se define una muestra reservada (consulte “Particiones ” en la página 211). Si especifica una variable para definir particiones, es importante que las características tengan distribuciones similares en todas las muestras reservadas, de entrenamiento o comprobación. Utilice, por ejemplo, el procedimiento Explorar para examinar las distribuciones en las particiones.

Ponderaciones de frecuencia. Este procedimiento ignora las ponderaciones de frecuencia.

Replicación de los resultados. El procedimiento utiliza la generación de números aleatorios durante la asignación aleatoria de particiones y pliegues de validación cruzada. Si desea duplicar los resultados de forma exacta, además de utilizar los mismos ajustes de procedimiento, defina una semilla para el Tornado de Mersenne (consulte “Particiones ” en la página 211), o utilice variables para definir particiones y pliegues de validación cruzada.

Para obtener un análisis de vecino más próximo

En los menús seleccione:

Analizar > **Clasificar** > **Vecino más próximo...**

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

1. Especifique una o más características que puedan constituir variables independientes o predictores en caso de haber un destino.

   Destino (opcional). Si no hay ningún destino (variable dependiente o respuesta) especificado, el procedimiento encontrará únicamente los *k* vecinos más próximos, sin realizar ninguna clasificación ni predicción.

   Características de escala de normalización. Las características normalizadas tienen el mismo rango de valores, lo que puede mejorar el rendimiento del algoritmo de estimación. Se utilizará la normalización ajustada [2*(x–min)/(max–min)]1. Los valores normalizados ajustados quedan comprendidos entre -1 y 1.

   Identificador de caso focal (opcional). Esto le permite marcar casos de especial interés. Por ejemplo, un investigador desea determinar si las puntuaciones de las pruebas de un distrito escolar (el caso focal) son comparables con las de distritos escolares similares. Utiliza un análisis de vecinos más próximos para encontrar los distritos escolares más parecidos con respecto a un conjunto dado de características. Después compara las puntuaciones de las pruebas del distrito escolar focal con las de los vecinos más próximos.

   Los casos focales también deben emplearse en estudios clínicos para seleccionar casos de control similares a los casos clínicos. Los casos focales se muestran en la tabla de *k* vecinos más próximos y distancias, el gráfico de espacio de características, el gráfico de homólogos y el mapa de cuadrantes. La información sobre casos focales se guarda en los archivos especificados en la pestaña Resultados.

   Los casos con un valor positivo en la variable especificada se tratan como casos focales. No es posible especificar una variable sin valores positivos.

Etiqueta de caso (opcional). Los casos se etiquetan utilizando estos valores en el gráfico de espacio de características, el gráfico de homólogos y el mapa de cuadrantes.

Campos con nivel de medición desconocido

La alerta de nivel de medición se muestra si el nivel de medición de una o más variables (campos) del conjunto de datos es desconocido. Como el nivel de medición afecta al cálculo de los resultados de este procedimiento, todas las variables deben tener un nivel de medición definido.

Explorar datos. Lee los datos del conjunto de datos activo y asigna el nivel de medición predefinido en cualquier campo con un nivel de medición desconocido. Si el conjunto de datos es grande, puede llevar algún tiempo.

Asignar manualmente. Abre un cuadro de diálogo que contiene todos los campos con un nivel de medición desconocido. Puede utilizar este cuadro de diálogo para asignar el nivel de medición a esos campos. También puede asignar un nivel de medición en la Vista de variables del Editor de datos.

Como el nivel de medición es importante para este procedimiento, no puede acceder al cuadro de diálogo para ejecutar este procedimiento hasta que se hayan definido todos los campos en el nivel de medición.

Vecinos

Número de vecinos más próximos (k) Especifique el número de vecinos más próximos. Tenga en cuenta que el uso de un número mayor de vecinos no implica que el modelo resultante sea más preciso.

Si se especifica un destino en la pestaña Variables, puede especificar un rango de valores y permitir que el procedimiento seleccione el "mejor" número de vecinos de ese rango. El método para determinar el número de vecinos más próximos depende de si se solicita la selección de características en la pestaña Características.

- Si la selección de características está activada, ésta se realizará para cada valor de *k* en el rango solicitado, y se seleccionará la *k* y el conjunto de funciones compañero con la menor tasa de error (o el menor error cuadrático si el destino es escala).
- Si la selección de características no está activada, se utilizará la validación cruzada de pliegue en *V* para seleccionar el “mejor” número de vecinos. Consulte la pestaña Partición para tener control sobre la asignación de pliegues.

Cálculo de distancias. Es la métrica utilizada para especificar la métrica de distancia empleada para medir la similitud de los casos.

- **Métrica euclídea.** La distancia entre dos casos, x e y, es la raíz cuadrada de la suma, sobre todas las dimensiones, de las diferencias cuadradas entre los valores de esos casos.
- **Métrica de bloques de ciudad.** La distancia entre dos casos es la suma, en todas las dimensiones, de las diferencias absolutas entre los valores de esos casos. También se conoce como la distancia de Manhattan.

Además, si se especifica un destino en la pestaña Variables, puede optar por ponderar características según su importancia normalizada a la hora de calcular distancias. La importancia que una característica tiene para un predictor se calcula en función de la relación entre la tasa de error o errores cuadráticos del modelo sin el predictor y la tasa de error o errores cuadráticos del modelo completo. La importancia normalizada se calcula volviendo a ponderar los valores de importancia de la característica para que sumen 1.

Predicciones del destino de escala. Si se especifica un destino de escala en la pestaña Variables, especificará si el valor predicho se calcula en función de la media o del valor de mediana de los vecinos más próximos.

Características

La pestaña Características le permite seleccionar y especificar opciones para la selección de características cuando se especifica un destino en la pestaña Variables. De forma predeterminada, todas las características se tienen en cuenta para la selección de características, pero es posible seleccionar un subconjunto de características para forzarlas en el modelo.

Criterio de parada. En cada paso, la característica cuya suma al modelo dé lugar al menor error (calculado como la tasa de error de un destino categórico y el error cuadrático de un destino de escala) se tiene en cuenta para su inclusión en el conjunto de modelos. La selección continúa hasta que se cumple la condición especificada.

- **Número de características especificadas.** El algoritmo añade un número fijo de características además de las forzadas en el modelo. Especifique un número entero positivo. Si se disminuyen los valores de número que se puede seleccionar se obtiene un modelo más reducido, lo que supone el riesgo de perder importantes características. Si se aumentan los valores de número que se puede seleccionar se incluirán todas las características importantes, pero se corre el riesgo de añadir características que aumenten el error del modelo.
- **Cambio mínimo de la tasa de errores absolutos.** El algoritmo se detiene cuando el cambio de la tasa de errores absolutos indica que el modelo no puede mejorarse más añadiendo nuevas características. Especifique un número positivo. Si se reducen los valores del cambio mínimo se incluirán más características, pero puede que se incluyan características que no añadan gran valor al modelo. Si se aumentan los valores del cambio mínimo se excluirán más características, pero puede que se pierdan características importantes para el modelo. El valor “óptimo” de cambio mínimo dependerá de sus datos y de la aplicación. Consulte el Registro de errores de selección de características en los resultados para poder evaluar qué características son más importantes. Consulte el tema “Registro de errores de selección de características ” en la página 215 para obtener más información.

Particiones

La pestaña Particiones le permite dividir el conjunto de datos en conjuntos de entrenamiento y reserva y, siempre que proceda, asignar casos a pliegues de validación cruzada.

Particiones de entrenamiento y reserva. Este grupo especifica el método de crear particiones en el conjunto de datos activo correspondientes a las muestras de entrenamiento y reserva. La **muestra de entrenamiento** comprende los registros de datos utilizados para entrenar el modelo de vecino más próximo; cierto porcentaje de casos del conjunto de datos debe asignarse a la muestra de entrenamiento para poder obtener un modelo. La **muestra reservada** es otro conjunto independiente de registros de datos que se utiliza para evaluar el modelo final; el error de la muestra reservada ofrece una estimación "sincera" de la capacidad predictora del modelo, ya que los casos reservados no se utilizan para crear el modelo.

- **Asignar casos a particiones aleatoriamente.** Especifique el porcentaje de casos que se asignarán a la muestra de entrenamiento. El resto se asignan a la muestra reservada.
- **Utilizar variable para asignar los casos.** Especifique una variable numérica que asigne cada caso del conjunto de datos activo a la muestra de entrenamiento o reserva. Los casos con un valor positivo de la variable se asignarán a la muestra de entrenamiento, los casos con un valor 0 o negativo se asignarán a la muestra reservada. Los casos con un valor perdido del sistema se excluirán del análisis. Todos los valores perdidos de usuario de la variable de partición se tratarán siempre como válidos.

Pliegues de validación cruzada. La validación cruzada de pliegue en *V* se utiliza para determinar el "mejor" número de vecinos. Por razones de rendimiento, no está disponible con la selección de características.

La validación cruzada divide la muestra en un número de submuestras o pliegues. A continuación, se generan los modelos de vecino más próximo, que no incluyen los datos de cada submuestra. El primer modelo se basa en todos los casos excepto los correspondientes al primer pliegue de la muestra; el segundo modelo se basa en todos los casos excepto los del segundo pliegue de la muestra y así sucesivamente. Para cada modelo se calcula el error aplicando el modelo a la submuestra que se excluyó al generarlo. El "mejor" número de vecinos más próximos será el que produzca el menor error entre los pliegues.

- **Asignar casos a pliegues aleatoriamente.** Especifique el número de pliegues que se utilizarán para la validación cruzada. El procedimiento asigna aleatoriamente casos a los pliegues, numerados de 1 a *V*, que es el número de pliegues.
- **Utilizar variable para asignar los casos.** Especifique una variable numérica que asigne cada caso del conjunto de datos activo a un pliegue. La variable debe ser numérica y tomar valores de 1 a *V*. Si

falta algún valor de este rango, y en cualquier división si los archivos de división están en vigor, esto provocará un error.

Definir semilla para tornado de Mersenne. Si se establece una semilla es posible replicar análisis. El uso de este control es parecido a establecer el tornado de Mersenne como generador activo y especificar un punto de inicio fijo en el cuadro de diálogo Generadores de números aleatorios, con la importante diferencia de que la definición de la semilla de este cuadro de diálogo mantendrá el estado actual del generador de números aleatorios y restaurará dicho estado cuando haya terminado el análisis.

Guardado

Nombres de las variables guardadas. La generación automática de nombres garantiza que conserva todo su trabajo. Los nombres personalizados le permiten descartar/reemplazar los resultados de las ejecuciones anteriores sin eliminar antes las variables guardadas en el Editor de datos.

Variables a guardar

- **Valor o categoría pronosticados.** Esta opción guarda el valor predicho para el destino de escala o la categoría predicha para un destino categórico.
- **Probabilidad predicha.** Esta opción guarda las probabilidades pronosticadas para un destino categórico. Para cada una de las primeras *n* categorías se guarda una variable diferente, donde *n* se especifica en el control **Máximo de categorías para guardar para un destino categórico**.
- **Variables de particiones de entrenamiento y reserva.** Si los casos se asignan aleatoriamente a las muestras de entrenamiento y reserva de la pestaña Particiones, esta opción guarda el valor de la partición (entrenamiento y reserva) a la que se ha asignado el caso.
- **Variable de pliegues de validación cruzada.** Si los casos se asignan aleatoriamente a los pliegues de validación cruzada de la pestaña Particiones, esta opción guarda el valor del pliegue al que se ha asignado el caso.

Resultado

Salida del visor

- **Resumen de procesamiento de casos.** Muestra la tabla de resumen de procesamiento de casos, que resume el número de casos incluidos y excluidos en el análisis, en total y por muestras de entrenamiento y reservadas.
- **Gráficos y tablas.** Muestra los resultados relacionados con los modelos, incluyendo tablas y gráficos. Las tablas de la vista de modelo incluyen los *k* vecinos más próximos y las distancias de casos focales, la clasificación de variables de respuesta categórica y un resumen de errores. El resultado gráfico de la vista de modelo incluye un registro de errores de selección, un gráfico de importancia de características, un gráfico de espacio de características, un gráfico de homólogos y un mapa de cuadrante. Consulte el tema “Vista de modelo ” en la página 213 para obtener más información.

Archivos

- **Exportar modelo a XML.** Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo. Esta opción no se encuentra disponible si se han definido archivos segmentados.
- **Exportar distancias entre casos focales y k vecinos más próximos.** En cada caso focal, se crea una variable distinta para cada uno de los *k* vecinos más próximos del caso focal (de la muestra de entrenamiento) y las *k* distancias más próximas correspondientes.

Opciones

Valores perdidos del usuario. Para que un caso se incluya en el análisis, las variables categóricas deben tener valores válidos para dicho caso. Estos controles permiten decidir si los valores perdidos del usuario se deben tratar como válidos entre las variables categóricas.

Los valores perdidos del sistema y perdidos para las variables de escala siempre se tratan como no válidos.

Vista de modelo

Cuando seleccione **Gráficos y tablas** en la pestaña Resultados, el procedimiento creará un objeto de modelo de vecino más próximo en el visor. Al activar (pulsando dos veces) este objeto se obtiene una vista interactiva del modelo. La vista de modelos tiene una ventana con dos paneles:

- El primer panel muestra una descripción general del modelo denominado vista principal.
- El segundo panel muestra uno de los dos tipos de vistas:

  Una vista de modelos auxiliar muestra más información sobre el modelo, pero no se centra en el propio modelo.

  Una vista enlazada es una vista que muestra detalles sobre una característica del modelo cuando el usuario desglosa parte de la vista principal.

De forma predeterminada, el primer panel muestra el espacio de características y el segundo muestra el gráfico de importancia de variables. Si el gráfico de importancia de variables no está disponible, es decir, si **Ponderar características por importancia** no se ha seleccionado en la pestaña Funciones, se mostrará la primera vista disponible en la lista desplegable Ver.

Cuando una vista no tiene ninguna información disponible, se desactiva este texto de elemento en la lista desplegable Ver.

Espacio de características

El gráfico de espacio de características es un gráfico interactivo del espacio de características (o un subespacio, si hay más de 3 características). Cada eje representa una característica del modelo, y la ubicación de los puntos del gráfico muestran los valores de dichas características para casos de las particiones de entrenamiento y reserva.

Llaves. Además los valores de las características, los puntos del gráfico indican otra información.

- La forma indica la partición a la que pertenece un punto, ya sea Entrenamiento o Reserva.
- El color y el sombreado de un punto indican el valor del destino de ese caso: cada valor de color diferente representa las categorías de un destino categórico y las sombras indican el rango de valores de un destino continuo. El valor indicado para la partición de entrenamiento es el valor observado, mientras que en el caso de la partición de reserva, representa el valor predicho. Si no se especifica ningún destino, esta clave no aparece.
- Los titulares más gruesos indican que un caso es focal. Los casos focales se muestran en relación con sus *k* vecinos más próximos.

Controles e interactividad. Una serie de controles del gráfico le permite explorar el espacio de características.

- Puede seleccionar qué subconjunto de características mostrar en el gráfico y modificar qué funciones se representan en las dimensiones.
- Los “casos focales” son simplemente puntos seleccionados en el gráfico del espacio de características. Si ha especificado una variable de caso focal, los puntos que representan los casos focales se seleccionarán inicialmente. Sin embargo, cualquier punto puede convertirse en un caso focal si lo selecciona. A la selección de puntos se aplican los controles “normales”, es decir, si pulsa en un punto éste se selecciona y se cancela la selección de todos los demás y si pulsa Control y el ratón sobre un punto éste se añadirá al conjunto de puntos seleccionados. Las vistas enlazadas, como el gráfico de homólogos, se actualizarán automáticamente en función de los casos seleccionados en el espacio de características.
- Puede modificar el número de vecinos más próximos (*k*) para mostrar casos focales.
- Al pasar el ratón sobre un punto del gráfico se mostrará una ayuda contextual con el valor de la etiqueta de caso o un número de caso si las etiquetas de caso no se definen, así como los valores de destino observados y pronosticados.
- Un botón “Restablecer” le permite devolver el espacio de características a su estado original.

Adición y eliminación de campos/variables

Puede añadir nuevos campos/variables al espacio de características o eliminar los que se visualizan en este momento.

Paleta de variables

Debe visualizar la paleta de variables antes de que pueda añadir y eliminar variables. Para visualizar la paleta de variables, el visor de modelos deberá estar en modo de edición y deberá seleccionarse un caso en el espacio de características.

1. Para poner el visor de modelos en modo de edición, elija en los menús:

   Ver > **Modalidad de edición**
2. Una vez en Modo Edición, pulse sobre cualquier caso del espacio de características.
3. Para visualizar la paleta de variables, elija en los menús:

   Ver > **Paletas** > **Variables**

   La paleta de variables enumera todas las variables del espacio de características. El icono junto al nombre de variable indica el nivel de medición de la variable.
4. Para cambiar temporalmente el nivel de medición de una variable, pulse con el botón derecho en la variable de la paleta de variables y seleccione una opción.

Zonas de variables

Las variables se añaden a "zonas" del espacio de características. Para visualizar las zonas, empiece arrastrando una variable desde la paleta de variables o seleccionando **Mostrar zonas**.

El espacio de características tiene zonas para los ejes *x*, *y* y *z*.

Desplazamiento de variables a zonas

Éstas son algunas reglas generales y sugerencias para desplazar variables a zonas:

- Para desplazar una variable a una zona, pulse y arrastre la variable desde la paleta de variables y suéltela en la zona. Si selecciona **Mostrar zonas**, también puede pulsar con el botón derecho en una zona y seleccionar una variable que desee añadir a la zona.
- Si arrastra una variable de la paleta de variables a una zona que ya esté ocupada por otra variable, la nueva variable sustituirá a la anterior.
- Si arrastra una variable de una zona a una zona que ya esté ocupada por otra variable, las variables intercambiarán posiciones.
- Si pulsa en la X de una zona, eliminará la variable de dicha zona.
- Si hay varios elementos gráficos en la visualización, cada elemento gráfico puede tener sus propias zonas de variables asociadas. Primero, seleccione el elemento gráfico.

Importancia de la variable

Normalmente, desea centrar sus esfuerzos de modelado en las variables que importan más y considera eliminar o ignorar las que importan menos. El gráfico de importancia de la variable le ayuda a hacerlo indicando la importancia relativa de cada variable en la estimación del modelo. Como las variables son relativas, la suma de los valores de todas las variables de la visualización es 1,0. La importancia de variable no está relacionada con la precisión del modelo. Sólo está relacionada con la importancia de cada variable para realizar una predicción, independientemente de si ésta es precisa o no.

Homólogos

Este gráfico muestra los casos focales y sus *k* vecinos más próximos en cada característica y en el destino. Está disponible si se selecciona un caso focal en el espacio de características.

Forma de enlace. El gráfico Homólogos se enlaza con el espacio de características de dos formas.

- Los casos seleccionados (focal) en el espacio de características se muestran en el gráfico Homólogos, juntos con sus *k* vecinos más próximos.
- El valor de *k* seleccionado en el espacio de características se utiliza en el gráfico Homólogos.

Distancias de vecinos más próximos

Esta tabla muestra los *k* vecinos más próximos y las distancias de casos focales únicamente. Está disponible si se especifica un identificador de caso focal en la pestaña Variables, y sólo muestra los casos focales identificados por esta variable.

Cada fila de:

- La columna **Caso focal** contiene el valor de la variable de etiqueta de caso del caso focal; si las etiquetas de caso no se definen, esta columna contendrá el número de caso del caso focal.
- La *i*ª columna del grupo Vecinos más próximos contiene el valor de la variable de etiqueta de caso del *i*º vecino más próximo al caso focal; si las etiquetas de caso no se definen, esta columna contendrá el número de caso del *i*º vecino más próximo al caso focal.
- La *i*ª columna del grupo Distancias más próximas contiene la distancia del *i*º vecino más próximo al caso focal.

Mapa de cuadrantes

Este gráfico muestra los casos focales y sus *k* vecinos más próximos en un diagrama de dispersión (o gráfico de puntos, dependiendo del nivel de medición del destino) con el destino en el eje *y* y una característica de escala en el eje *x*, panelado por características. Está disponible si hay un destino y se selecciona un caso focal en el Espacio de características.

- Se dibujan líneas de referencia para las variables continuas en las medias variables en la partición de entrenamiento.

Registro de errores de selección de características

Señala en la vista de gráfico el error (la tasa de error o de error cuadrático, dependiendo del nivel de medición del destino) en el eje *y* para el modelo con la característica enumerada en el eje *x* (además de todas las características a la izquierda del eje *x*). Este gráfico está disponible si hay un destino y la selección de características está activada.

Registro de errores de selección de k

Señala en la vista de gráfico el error (la tasa de error o de error cuadrático, dependiendo del nivel de medición del destino) en el eje *y* para el modelo con el número de vecinos más próximos (*k*) en el eje *x*. Este gráfico está disponible si hay un destino y la selección de *k* está activada.

Registro de errores de selección de características y k

Estos son los gráficos de selección de características (consulte “Registro de errores de selección de características ” en la página 215), con paneles por *k*. Este gráfico está disponible si hay un destino y *k* y la selección de características están en vigor.

Tabla de clasificación

Esta tabla muestra la clasificación cruzada de los valores observados en comparación con los valores pronosticados del destino, en función de la partición. Está disponible si hay un destino y es categórico.

- La fila **(Perdidos)** de la partición de reserva contiene casos de reserva con los valores perdidos en el destino. Estos casos contribuyen a los valores de Muestra reservada: Valores de Porcentaje global, pero no a los valores de Porcentaje correcto.

Resumen de error

Esta tabla está disponible si hay una variable objetivo. Muestra el error asociado con el modelo, la suma de cuadrados de un destino continuo y la tasa de error (100%, porcentaje global correcto) de un destino categórico.

Análisis discriminante

El análisis discriminante crea un modelo predictivo para la pertenencia al grupo. El modelo está compuesto por una función discriminante (o, para más de dos grupos, un conjunto de funciones discriminantes) basada en combinaciones lineales de las variables predictoras que proporcionan la mejor discriminación posible entre los grupos. Las funciones se generan a partir de una muestra de casos para los que se conoce el grupo de pertenencia; posteriormente, las funciones pueden ser aplicadas a nuevos casos que dispongan de mediciones para las variables predictoras pero de los que se desconozca el grupo de pertenencia.

Nota: la variable de agrupación puede tener más de dos valores. Los códigos de la variable de agrupación han de ser números enteros y es necesario especificar sus valores máximo y mínimo. Los casos con valores fuera de estos límites se excluyen del análisis.

Ejemplo. Por término medio, las personas de los países de zonas templadas consumen más calorías por día que las de los trópicos, y una proporción mayor de la población de las zonas templadas vive en núcleos urbanos. Un investigador desea combinar esta información en una función para determinar cómo de bien un individuo es capaz de discriminar entre los dos grupos de países. El investigador considera además que el tamaño de la población y la información económica también pueden ser importantes. El análisis discriminante permite estimar los coeficientes de la función discriminante lineal, que tiene el aspecto de la parte derecha de una ecuación de regresión lineal múltiple. Es decir, utilizando los coeficientes *a*, *b*, *c* y *d*, la función es:

```
D = a * climate + b * urban + c * population + d * gross domestic product per capita
```

Si estas variables resultan útiles para discriminar entre las dos zonas climáticas, los valores de *D* serán diferentes para los países templados y para los tropicales. Si se utiliza un método de selección de variables por pasos, quizás no se necesite incluir las cuatro variables en la función.

Estadísticas. Para cada variable: medias, desviaciones estándar, ANOVA univariado. Para cada análisis: *M* de Box, matriz de correlaciones intra-grupos, matriz de covarianzas intra-grupos, matriz de covarianzas de los grupos separados, matriz de covarianzas total. Para cada función discriminante canónica: autovalores, porcentaje de varianza, correlación canónica, lambda de Wilks, chi-cuadrado. Para cada paso: probabilidades previas, coeficientes de la función de Fisher, coeficientes de función no tipificados, lambda de Wilks para cada función canónica.

Análisis discriminante: Consideraciones sobre los datos

Datos. La variable de agrupación debe tener un número limitado de categorías distintas, codificadas como números enteros. Las variables independientes que sean nominales deben ser recodificadas a variables auxiliares o de contraste.

Supuestos. Los casos deben ser independientes. Las variables predictoras deben tener una distribución normal multivariada y las matrices de varianzas-covarianzas intra-grupos deben ser iguales en todos los grupos. Se asume que la pertenencia al grupo es mutuamente exclusiva (es decir, ningún caso pertenece a más de un grupo) y exhaustiva de modo colectivo (es decir, todos los casos son miembros de un grupo). El procedimiento es más efectivo cuando la pertenencia al grupo es una variable verdaderamente categórica; si la pertenencia al grupo se basa en los valores de una variable continua (por ejemplo, un cociente de inteligencia alto respecto a uno bajo), considere el uso de la regresión lineal para aprovechar la información más rica ofrecida por la propia variable continua.

Para obtener un análisis discriminante

1. En los menús seleccione:

   Analizar > **Clasificar** > **Discriminante...**

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione una variable de agrupación con valores enteros y pulse en **Definir rango** para especificar las categorías de interés.
3. Seleccione las variables independientes o predictoras. (Si la variable de agrupación no tiene valores enteros, la opción Recodificación automática en el menú Transformar creará una variable que los tenga).
4. Seleccione el método de introducción de las variables independientes.
   - **Introducir independientes juntas.** Introducir simultáneamente todas las variables independientes que satisfacen el criterio de tolerancia.
   - **Usar método de inclusión por pasos.** Utiliza el análisis por pasos para controlar la entrada y la eliminación de variables.
5. Si lo desea, seleccione casos mediante una variable de selección.

Análisis discriminante: Definir rango

Especifique los valores mínimo y máximo de la variable de agrupación para el análisis. Los casos con valores fuera de este rango no se utilizan en el análisis discriminante, pero sí se clasifican en uno de los grupos existentes a partir de los resultados que obtengan en el análisis. Los valores mínimo y máximo deben ser números enteros.

Análisis discriminante: Seleccionar casos

Para seleccionar casos para el análisis:

1. En el cuadro de diálogo Análisis discriminante, seleccione una variable de selección.
2. Pulse en **Valor** para introducir un número entero como valor de selección.

Sólo se utilizan los casos con el valor especificado en la variable de selección para derivar las funciones discriminantes. Tanto para los casos seleccionados como para los no seleccionados se generan resultados de clasificaciones y estadísticos. Este proceso ofrece un mecanismo para clasificar casos nuevos basados en datos previos o para dividir los datos en subconjuntos de contraste y comprobación para realizar procedimientos de validación en el modelo generado.

Análisis discriminante: Estadísticos

Descriptivos. Las opciones disponibles son: Medias (que incluye las desviaciones estándar), ANOVAs univariados y prueba *M* de Box.

- *Medias*. Muestra la media y desviación estándar totales y las medias y desviaciones estándar de grupo, para las variables independientes.
- *ANOVAs univariados*. Realiza un análisis de varianza de un factor sobre la igualdad de las medias de grupo para cada variable independiente.
- *M de Box*. Una prueba para la igualdad de las matrices de covarianza del grupo. Para tamaños de muestras suficientemente grandes, un valor de p no significativo quiere decir que no hay suficiente evidencia de que las varianzas sean diferentes. Esta prueba es sensible a las desviaciones de la normalidad multivariada.

Coeficientes de la función. Las opciones disponibles son: Coeficientes de clasificación de Fisher y Coeficientes no estandarizados.

- *De Fisher*. Muestra los coeficientes de la función de clasificación de Fisher que pueden utilizarse directamente para la clasificación. Se obtiene un conjunto de coeficientes para cada grupo, y se asigna un caso al grupo para el que tiene una mayor puntuación discriminante (valor de función de clasificación).
- *No estandarizados*. Muestra los coeficientes de la función discriminante sin estandarizar.

Matrices. Las matrices de coeficientes disponibles para las variables independientes son las de: Correlación intra-grupos, Covarianza intra-grupos, Covarianza de grupos separados y Covarianza total.

- *Correlación intra-grupos*. Muestra la matriz de correlaciones intra-grupos combinada, que se obtiene de promediar las matrices de covarianza individuales para todos los grupos antes de calcular las correlaciones.
- *Covarianza intra-grupos*. Muestra la matriz de covarianza intra-grupos agrupada, la cual puede diferir de la matriz de covarianza total. La matriz se obtiene de promediar, para todos los grupos, las matrices de covarianza individuales.
- *Covarianza de grupos separados*. Muestra las matrices de covarianza de cada grupo por separado.
- *Covarianza total*. Muestra la matriz de covarianza para todos los casos, como si fueran una única muestra.

Análisis discriminante: Método de inclusión por pasos

Método. Seleccione el estadístico que se va a utilizar para introducir o eliminar nuevas variables. Las alternativas disponibles son la lambda de Wilks, la varianza inexplicable, la distancia Mahalanobis, la relación más pequeña *F* y la *V*de Rao. Con el *V*de Rao, puede especificar el aumento mínimo en *V* para que entre una variable.

- *Lambda de Wilks*. Un método de selección de variables para el análisis discriminante por pasos que elige variables para la entrada en la ecuación en función de cuánto bajan la lambda de Wilks. En cada paso, se entra la variable que minimiza la lambda de Wilks global.
- *Varianza inexplicable*. En cada paso se introduce la variable que minimiza la suma de la variación no explicada entre los grupos.
- *Distancia de Mahalanobis*. Medida de cuánto difieren los valores de un caso en las variables independientes del promedio de todos los casos. Una distancia de Mahalanobis grande identifica un caso que tenga valores extremos en una o más de las variables independientes.
- *Relación F más pequeña*. Método para la selección de variables en los análisis por pasos que se basa en maximizar la razón F, calculada a partir de la distancia de Mahalanobis entre los grupos.
- *V de Rao*. Medida de las diferencias entre medias de grupo. También se denomina la traza de Lawley-Hotelling. En cada paso, se incluye la variable que maximiza el incremento de la V de Rao. Después de seleccionar esta opción, introduzca el valor mínimo que debe tener una variable para poder incluirse en el análisis.

Criterios. Las alternativas disponibles son **Utilizar valor F** y **Utilizar probabilidad de F**. Especifique valores para incorporar y eliminar variables.

- *Utilizar valor F*. Se introduce una variable en el modelo si su valor F es mayor que el valor de Entrada y se elimina si el valor F es menor que el valor de Eliminación. La entrada debe ser mayor que la eliminación y ambos valores deben ser positivos. Para introducir más variables en el modelo, disminuya el valor de entrada. Para eliminar más variables del modelo, eleve el valor de eliminación.
- *Usar probabilidad de F*. Una variable se introduce en el modelo si el nivel de significación de su valor F es menor que el valor de Entrada y se elimina si el nivel de significación es mayor que el valor de Eliminación. La entrada debe ser menor que la eliminación y ambos valores deben ser positivos. Para introducir más variables en el modelo, aumente el valor de entrada. Para eliminar más variables del modelo, disminuya el valor de eliminación.

Representación. Resumen de los pasos muestra las estadísticas para todas las variables después de cada paso; **F para las distancias por pares** muestra una matriz de relaciones *F* por pares para cada par de grupos.

Análisis discriminante: Clasificar

Probabilidades previas. Esta opción determina si se corrigen los coeficientes de clasificación teniendo en cuenta la información previa sobre la pertenencia a los grupos.

- **Todos los grupos iguales.** Se suponen probabilidades previas iguales para todos los grupos. Esta opción no tiene ningún efecto sobre los coeficientes.
- **Calcular según tamaños de grupos.** Los tamaños de los grupos observados de la muestra determinan las probabilidades previas de la pertenencia a los grupos. Por ejemplo, si el 50% de las observaciones incluidas en el análisis corresponden al primer grupo, el 25% al segundo y el 25% al tercero, se corregirán los coeficientes de clasificación para aumentar la probabilidad de la pertenencia al primer grupo respecto a los otros dos.

Representación. Las opciones de presentación disponibles son: Resultados por casos, Tabla de resumen y Clasificación dejando uno fuera.

- *Resultados por casos*. Se muestran, para cada caso, los códigos del grupo real de pertenencia, el grupo pronosticado, las probabilidades posteriores y las puntuaciones discriminantes.
- *Tabla de resumen*. Número de casos correcta e incorrectamente asignados a cada uno de los grupos, basándose en el análisis discriminante. En ocasiones se denomina "Matriz de Confusión".
- *Clasificación dejando uno fuera*. Se clasifica cada caso del análisis por las funciones derivadas de todos los casos, excepto el propio caso. También se conoce como método U.

Reemplazar los valores perdidos con la media. Seleccione esta opción para sustituir la media de una variable independiente para un valor perdido sólo durante la fase de clasificación.

Usar matriz de covarianzas. Existe la opción de clasificar los casos utilizando una matriz de covarianzas intra-grupos o una matriz de covarianzas de los grupos separados.

- *Intra-grupos*. Se utiliza la matriz de covarianza intra-grupos combinada para clasificar los casos.
- *Grupos separados*. Para la clasificación se utilizan las matrices de covarianza de los grupos separados. Dado que la clasificación se basa en las funciones discriminantes y no en las variables originales, esta opción no siempre es equivalente a la discriminación cuadrática.

Diagramas. Las opciones de gráficos disponibles son: Grupos combinados, Grupos separados y Mapa territorial.

- *Grupos combinados*. Crea un diagrama de dispersión de todos los grupos de los dos primeros valores de función discriminante. Si sólo hay una función, en su lugar se muestra un histograma.
- *Grupos separados*. Crea diagramas de dispersión de grupos separados de los dos primeros valores de función discriminante. Si sólo hay una función, en su lugar se muestra un histograma.
- *Mapa territorial*. Un gráfico de los límites utilizados para clasificar los casos en grupos basándose en valores de función. Los números corresponden a los grupos en los que se clasifican los casos. La media de cada grupo se indica mediante un asterisco situado dentro de sus fronteras. No se mostrará el mapa si sólo hay una función discriminante.

Análisis discriminante: Guardar

Es posible añadir variables nuevas al archivo de datos activo. Las opciones disponibles son las de grupo de pertenencia pronosticado (una única variable), puntuaciones discriminantes (una variable para cada función discriminante en la solución) y probabilidades de pertenencia al grupo según las puntuaciones discriminantes (una variable para cada grupo).

También se puede exportar información del modelo al archivo especificado en formato XML. Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo.

Características adicionales del comando DISCRIMINANT

La sintaxis de comandos también le permite:

- Realizar varios análisis discriminantes (con un comando) y controlar el orden en el que se introducen las variables (mediante el subcomando ANALYSIS).
- Especificar probabilidades previas para la clasificación (mediante el subcomando PRIORS).
- Mostrar matrices de estructura y de configuración rotadas (mediante el subcomando ROTATE).

- Limitar el número de funciones discriminantes extraídas (mediante el subcomando FUNCTIONS).
- Restringir la clasificación a los casos que están seleccionados (o no seleccionados) para el análisis (mediante el subcomando SELECT).
- Leer y analizar una matriz de correlaciones (mediante el subcomando MATRIX).
- Escribir una matriz de correlaciones para su análisis posterior (mediante el subcomando MATRIX).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Análisis factorial

El análisis factorial intenta identificar variables subyacentes, o **factores**, que expliquen la configuración de las correlaciones dentro de un conjunto de variables observadas. El análisis factorial se suele utilizar en la reducción de los datos para identificar un pequeño número de factores que explique la mayoría de la varianza observada en un número mayor de variables manifiestas. También puede utilizarse para generar hipótesis relacionadas con los mecanismos causales o para inspeccionar las variables para análisis subsiguientes (por ejemplo, para identificar la colinealidad antes de realizar un análisis de regresión lineal).

El procedimiento de análisis factorial ofrece un alto grado de flexibilidad:

- Existen siete métodos de extracción factorial disponibles.
- Existen cinco métodos de rotación disponibles, entre ellos el oblimin directo y el promax para rotaciones no ortogonales.
- Existen tres métodos disponibles para calcular las puntuaciones factoriales; y las puntuaciones pueden guardarse como variables para análisis adicionales.

Ejemplo. ¿Qué actitudes subyacentes hacen que las personas respondan a las preguntas de una encuesta política de la manera en que lo hacen? Examinando las correlaciones entre los elementos de la encuesta se deduce que hay una superposición significativa entre los diversos subgrupos de elementos (las preguntas sobre los impuestos tienden a estar correlacionadas entre sí, las preguntas sobre temas militares también están correlacionadas entre sí, y así sucesivamente). Con el análisis factorial, se puede investigar el número de factores subyacentes y, en muchos casos, identificar lo que los factores representan conceptualmente. Adicionalmente, se pueden calcular las puntuaciones factoriales para cada encuestado, que pueden utilizarse en análisis subsiguientes. Por ejemplo, es posible generar un modelo de regresión logística para predecir el comportamiento de voto basándose en las puntuaciones factoriales.

Estadísticas. Para cada variable: número de casos válidos, media y desviación estándar. Para cada análisis factorial: matriz de correlaciones de variables, incluidos niveles de significación, determinante, inversa; matriz de correlaciones reproducida, que incluye anti-imagen; solución inicial (comunalidades, autovalores y porcentaje de varianza explicada); KMO (medida de la adecuación muestral de Kaiser-Meyer-Olkin) y prueba de esfericidad de Bartlett; solución sin rotar, que incluye cargas factoriales, comunalidades y autovalores; y solución rotada, que incluye la matriz de configuración rotada y la matriz de transformación. Para rotaciones oblicuas: las matrices de estructura y de configuración rotadas; matriz de coeficientes para el cálculo de las puntuaciones factoriales y matriz de covarianzas entre los factores. Gráficos: gráfico de sedimentación y gráfico de las cargas de los dos o tres primeros factores.

Análisis factorial: Consideraciones sobre los datos

Datos. Las variables deben ser cuantitativas a nivel de *intervalo* o de *razón*. Los datos categóricos (como la religión o el país de origen) no son adecuados para el análisis factorial. Los datos para los cuales razonablemente se pueden calcular los coeficientes de correlación de Pearson, deberían ser adecuados para el análisis factorial.

Supuestos. Los datos deben tener una distribución normal bivariada para cada pareja de variables y las observaciones deben ser independientes. El modelo de análisis factorial especifica que las variables vienen determinadas por los factores comunes (los factores estimados por el modelo) y por factores exclusivos (los cuales no se superponen entre las distintas variables observadas); las estimaciones calculadas se basan en el supuesto de que ningún factor único está correlacionado con los demás, ni con los factores comunes.

Para obtener un análisis factorial

1. En los menús seleccione:

   Analizar > **Reducción de dimensiones** > **Factor...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione las variables para el análisis factorial.

Selección de casos en el análisis factorial

Para seleccionar casos para el análisis:

1. Seleccione una variable de selección.
2. Pulse en **Valor** para introducir un número entero como valor de selección.

En el análisis factorial, sólo se usarán los casos con ese valor para la variable de selección.

Análisis factorial: Descriptivos

Estadísticas Descriptivos univariados incluye la media, la desviación estándar y el número de casos válidos para cada variable. La **solución inicial** muestra las comunalidades iniciales, los autovalores y el porcentaje de varianza explicada.

Matriz de correlaciones. Las opciones disponibles son: coeficientes, niveles de significación, determinante, inversa, reproducida, anti-imagen y KMO y prueba de esfericidad de Bartlett.

- *Prueba de esfericidad de KMO y Bartlett*. La medida Kaiser-Meyer-Olkin de la adecuación del muestreo comprueba si las correlaciones parciales entre las variables son pequeñas. La prueba de esfericidad de Bartlett contrasta si la matriz de correlaciones es una matriz de identidad, que indicaría que el modelo factorial es inadecuado.
- *Reproducidos*. La matriz de correlaciones estimada de la solución factorial. También se muestran las correlaciones de residuos(la diferencia entre la correlación observada y la estimada).
- *Anti-imagen*. La matriz de correlaciones anti-imagen contiene los negativos de los coeficientes de correlación parcial, y la matriz de covarianzas anti-imagen contiene los negativos de las covarianzas parciales. En un buen modelo factorial la mayoría de los elementos no diagonales deben ser pequeños. En la diagonal de la matriz de correlaciones anti-imagen se muestra la medida de adecuación muestral para esa variable.

Análisis factorial: Extracción

Método. Permite especificar el método de extracción factorial. Los métodos disponibles son: Componentes principales, Mínimos cuadrados no ponderados, Mínimos cuadrados generalizados, Máxima verosimilitud, factorización de Ejes principales, factorización Alfa y factorización Imagen.

- *Análisis de componentes principales*. Método de extracción factorial utilizado para formar combinaciones lineales no correlacionadas de las variables observadas. El primer componente tiene la varianza máxima. Las componentes sucesivas explican progresivamente proporciones menores de la varianza y no están correlacionadas unas con otras. El análisis principal de las componentes se utiliza para obtener la solución factorial inicial. No se puede utilizar cuando una matriz de correlaciones es singular.
- *Método de mínimos cuadrados no ponderados*. Método de extracción de factores que minimiza la suma de los cuadrados de las diferencias entre las matrices de correlación observada y reproducida, ignorando las diagonales.
- *Método de Mínimos cuadrados generalizados*. Método de extracción factorial que minimiza la suma de las diferencias cuadráticas entre las matrices de correlación observadas y reproducidas. Las correlaciones se ponderan por el inverso de su exclusividad, de manera que las variables que tengan un valor alto de exclusividad reciban una ponderación menor que aquéllas que tengan un valor bajo de exclusividad.

- *Método de máxima verosimilitud*. Método de extracción factorial que genera estimaciones de parámetros que tienen más probabilidades de haber producido la matriz de correlaciones observada si la muestra procede de una distribución normal multivariante. Las correlaciones se ponderan por el inverso de la exclusividad de las variables, y se emplea un algoritmo iterativo.
- *Factorización de ejes principales*. Método de extracción de factores de la matriz de correlaciones original, con coeficientes de correlación múltiple al cuadrado colocados en la diagonal como estimaciones iniciales de las comunalidades. Las cargas factoriales resultantes se utilizan para estimar de nuevo las comunalidades que reemplazan a las estimaciones previas de comunalidad en la diagonal. Las iteraciones continúan hasta que el cambio en las comunalidades, de una iteración a la siguiente, satisfaga el criterio de convergencia para la extracción.
- *Factorización Alfa*. Método de extracción factorial que considera que las variables del análisis son una muestra del universo de variables potenciales. Este método maximiza el Alfa de Cronbach para los factores.
- *Factorización de imágenes*. Método de extracción de factores desarrollado por Guttman y basado en la teoría de imágenes. La parte común de una variable, llamada la imagen parcial, se define como su regresión lineal sobre las restantes variables, en lugar de ser una función de los factores hipotéticos.

Analizar. Permite especificar o una matriz de correlaciones o una matriz de covarianzas.

- **Matriz de correlaciones.** Es útil si las variables de su análisis se miden sobre escalas distintas.
- **Matriz de covarianzas.** Es útil si se desea aplicar el análisis factorial a varios grupos con distintas varianzas para cada variable.

Extraer. Se pueden retener todos los factores cuyos autovalores excedan un valor especificado o retener un número específico de factores.

Representación. Permite solicitar la solución factorial sin rotar y el gráfico de sedimentación de los autovalores.

- *Solución factorial sin rotar*. Muestra las cargas factoriales sin rotar (la matriz de configuración factorial), las comunalidades y los autovalores de la solución factorial.
- *Diagrama de sedimentación*. Un gráfico de la varianza que esta asociado a cada factor. Este gráfico se utiliza para determinar cuántos factores se deben retenerse. Típicamente el gráfico muestra una clara ruptura entre la pronunciada inclinación de los factores más importantes y el descenso gradual de los restantes (los sedimentos).

Nº máximo de iteraciones para convergencia. Permite especificar el número máximo de pasos que el algoritmo puede seguir para estimar la solución.

Análisis factorial: Rotación

Método. Le permite seleccionar el método de rotación de factores y si desea aplicar la normalización Kaiser. Los métodos disponibles son: varimax, equamax, quartimax, oblimin directo y promax.

- *Método Varimax*. Método de rotación ortogonal que minimiza el número de variables que tienen cargas altas en cada factor. Simplifica la interpretación de los factores.
- *Método Oblimin directa*.Método para la rotación oblicua (no ortogonal). Si delta es igual a cero (el valor predeterminado) las soluciones son las más oblicuas. A medida que delta se va haciendo más negativo, los factores son menos oblicuos. Para anular el valor predeterminado 0 para delta, introduzca un número menor o igual que 0,8.
- *Método quartimax*. Método de rotación que minimiza el número de factores necesarios para explicar cada variable. Simplifica la interpretación de las variables observadas.
- *Método Equamax* . Método de rotación que es una combinación del método varimax, que simplifica los factores, y el método quartimax, que simplifica las variables. Se minimiza tanto el número de variables que saturan alto en un factor como el número de factores necesarios para explicar una variable.
- *Rotación Promax*. Una rotación oblicua, que permite correlacionar factores. Esta rotación se puede calcular más rápidamente que una rotación oblimin directa, por lo que es útil para conjuntos de datos grandes.

- *Aplicar normalización Kaiser*. De forma predeterminada, le permite aplicar la normalización Kaiser cuando se especifica una rotación.

Representación. Permite incluir los resultados de la solución rotada, así como los gráficos de las cargas para los dos o tres primeros factores.

- *Solución rotada*. Se debe seleccionar un método de rotación para obtener una solución rotada. Para las rotaciones ortogonales, se muestran la matriz de configuración rotada y la matriz de transformación de factor. Para las rotaciones oblicuas, se muestran las matrices de correlaciones de factor, estructura y patrón.
- *Diagrama de las cargas factoriales*. Gráfico de carga factorial tridimensional de los tres primeros factores. En una solución de dos factores, se representa un diagrama bidimensional. Si sólo se extrae un factor no se muestra el gráfico. Si se solicita la rotación, los diagramas representan las soluciones rotadas.

Nº máximo de iteraciones para convergencia. Permite especificar el número máximo de pasos que el algoritmo puede seguir para llevar a cabo la rotación.

Análisis factorial: Puntuaciones factoriales

Guardar como variables. Crea una nueva variable para cada factor en la solución final.

Método. Los métodos alternativos para calcular las puntuaciones factoriales son: regresión, Bartlett, y Anderson-Rubin.

- *Método de regresión*. Método para estimar los coeficientes de puntuación factorial. Las puntuaciones que se producen tienen una media de 0 y una varianza igual al cuadrado de la correlación múltiple entre las puntuaciones factoriales estimadas y los valores factoriales verdaderos. Las puntuaciones puede correlacionarse incluso si los factores son ortogonales.
- *Puntuación de Bartlett*. Método de estimación de los coeficientes de puntuación factorial. Las puntuaciones que se generan tienen una media de 0. La suma de cuadrados de los factores únicos sobre el rango de variables se minimiza.
- *Método de Anderson-Rubin*. Un método de estimación de los coeficientes de puntuación factorial; una modificación del método Bartlett que garantiza la ortogonalidad de los factores estimados. Las puntuaciones resultantes tienen una media 0, una desviación estándar de 1 y no correlacionan entre sí.

Mostrar matriz de coeficientes de las puntuaciones factoriales. Muestra los coeficientes por los cuales se multiplican las variables para obtener puntuaciones factoriales. También muestra las correlaciones entre las puntuaciones factoriales.

Análisis factorial: Opciones

Valores perdidos. Permite especificar el tratamiento que reciben los valores perdidos. Las selecciones disponibles son: Excluir casos *según lista*, Excluir casos *según pareja* y Reemplazar por la media.

Formato de presentación de los coeficientes. Permite controlar aspectos de las matrices de resultados. Los coeficientes se ordenan por tamaño y se suprimen aquellos cuyos valores absolutos sean menores que el valor especificado.

Características adicionales del comando FACTOR

La sintaxis de comandos también le permite:

- Especificar los criterios de convergencia para la iteración durante la extracción y la rotación.
- Especificar gráficos factoriales rotados individuales.
- Especificar el número de puntuaciones factoriales que se van a guardar.
- Especificar valores diagonales para el método de factorización del eje principal.
- Escribir matrices de correlación o matrices de carga factorial en el disco para su análisis posterior.
- Leer y analizar matrices de correlación o matrices de carga factorial.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Selección de procedimientos para la agrupación en clústeres

Los análisis de clústeres se pueden realizar mediante los procedimientos de análisis de clústeres en dos fases, jerárquico o de K-medias. Cada uno de estos procedimientos emplea un algoritmo distinto en la creación de clústeres y contiene opciones que no están disponibles en los otros.

Análisis de clústeres en dos fases. En algunas aplicaciones, se puede seleccionar como método el procedimiento Análisis de clústeres en dos fases. Ofrece una serie de características exclusivas que se detallan a continuación:

- Selección automática del número más apropiado de clústeres y medidas para la selección de los distintos modelos de clúster.
- Posibilidad de crear modelos de clúster basados al mismo tiempo en variables categóricas y continuas.
- Posibilidad de guardar el modelo de clúster en un archivo XML externo y, a continuación, leer el archivo y actualizar el modelo de clúster con datos más recientes.

Asimismo, el procedimiento Análisis de clústeres en dos fases puede analizar archivos de datos grandes.

Análisis de clústeres jerárquico. El uso del procedimiento Análisis de clústeres jerárquico se limita a archivos de datos más pequeños (cientos de objetos por agrupar en clústeres) y ofrece una serie de características exclusivas que se detallan a continuación:

- Posibilidad de agrupar en clústeres casos o variables.
- Posibilidad de calcular un rango de soluciones posibles y guardar los clústeres de pertenencia para cada una de dichas soluciones.
- Distintos métodos de formación de clústeres, transformación de variables y medida de disimilaridad entre clústeres.

Siempre que todas las variables sean del mismo tipo, el procedimiento Análisis de clústeres jerárquico podrá analizar variables de intervalo (continuas), de recuento o binarias.

Análisis de clústeres de K-medias. El uso del procedimiento Análisis de clústeres de K-medias se limita a datos continuos y requiere que el usuario especifique previamente el número de clústeres y ofrece una serie de características exclusivas que se detallan a continuación:

- Posibilidad de guardar las distancias desde los centros de los clústeres hasta los distintos objetos.
- Posibilidad de leer los centros de los clústeres iniciales y guardar los centros de los clústeres finales desde un archivo IBM SPSS Estadística externo.

Asimismo, el procedimiento Análisis de clústeres de K-medias puede analizar archivos de datos grandes.

Análisis de clústeres bietápicos

El procedimiento Análisis de clústeres en dos fases es una herramienta de exploración diseñada para descubrir las agrupaciones naturales (o clústeres) de un conjunto de datos que, de otra manera, no sería posible detectar. El algoritmo que emplea este procedimiento incluye varias atractivas características que lo hacen diferente de las técnicas de agrupación en clústeres tradicionales:

- **Tratamiento de variables categóricas y continuas.** Al suponer que las variables son independientes, es posible aplicar una distribución normal multinomial conjunta en las variables continuas y categóricas.
- **Selección automática del número de clústeres.** Mediante la comparación de los valores de un criterio de selección del modelo para diferentes soluciones de agrupación en clústeres, el procedimiento puede determinar automáticamente el número óptimo de clústeres.
- **Escalabilidad.** Mediante la construcción de un árbol de características de clústeres (CF) que resume los registros, el algoritmo en dos fases puede analizar archivos de datos de gran tamaño.

Ejemplo. Las empresas minoristas y de productos de consumo aplican regularmente técnicas de agrupación a los datos que describen los hábitos de compra de sus clientes, el sexo, la edad, el nivel de ingresos, etc. Estas empresas adaptan sus estrategias de marketing y desarrollo de productos a cada grupo de consumidores para aumentar las ventas y el nivel de fidelidad a la marca.

Medida de distancia. Esta opción determina cómo se calcula la similaridad entre dos clústeres.

- **Log-verosimilitud.** La medida de la verosimilitud realiza una distribución de probabilidad entre las variables. Las variables continuas se supone que tienen una distribución normal, mientras que las variables categóricas se supone que son multinomiales. Se supone que todas las variables son independientes.
- **Euclídea.** La medida euclídea es la distancia según una "línea recta" entre dos clústeres. Sólo se puede utilizar cuando todas las variables son continuas.

Número de clústeres. Esta opción permite especificar cómo se va a determinar el número de clústeres.

- **Determinar automáticamente.** El procedimiento determinará automáticamente el número "óptimo" de clústeres, utilizando el criterio especificado en el grupo Criterio de agrupación en clústeres. Si lo desea, introduzca un entero positivo para especificar el número máximo de clústeres que el procedimiento debe tener en cuenta.
- **Especificar número fijo.** Permite fijar el número de clústeres de la solución. Introduzca un número entero positivo.

Recuento de variables continuas. Este grupo proporciona un resumen de las especificaciones acerca de la tipificación de variables continuas realizadas en el cuadro de diálogo Opciones. Consulte el tema “Opciones del análisis de clústeres en dos fases” en la página 226 para obtener más información.

Criterio de agrupación en clústeres. Esta opción determina cómo el algoritmo de agrupación en clústeres determina el número de clústeres. Se puede especificar tanto el criterio de información bayesiano (BIC) como el criterio de información de Akaike (AIC).

Consideraciones sobre los datos para el análisis de clústeres en dos fases

Datos. Este procedimiento trabaja tanto con variables continuas como categóricas. Los casos representan los objetos a agrupar en clústeres y las variables representan los atributos en los que se va a basar la agrupación en clústeres.

Orden de casos. Observe que el árbol de características de clústeres y la solución final pueden depender del orden de los casos. Para minimizar los efectos del orden, ordene los casos aleatoriamente. Puede que desee obtener varias soluciones distintas con los casos ordenados en distintos órdenes aleatorios para comprobar la estabilidad de una solución determinada. En situaciones en que esto resulta difícil debido a unos tamaños de archivo demasiado grandes, se pueden sustituir varias ejecuciones por una muestra de casos ordenados con distintos órdenes aleatorios.

Supuestos. La medida de la distancia de la verosimilitud supone que las variables del modelo de clúster son independientes. Además, se supone que cada variable continua tiene una distribución normal (de Gauss) y que cada variable categórica tiene una distribución multinomial. Las comprobaciones empíricas internas indican que este procedimiento es bastante robusto frente a las violaciones tanto del supuesto de independencia como de las distribuciones, pero aún así es preciso tener en cuenta hasta qué punto se cumplen estos supuestos.

Utilice el procedimiento de correlaciones bivariadas para comprobar la independencia de dos variables continuas. Utilice el procedimiento de tablas cruzadas para comprobar la independencia de dos variables categóricas. Utilice el procedimiento de medias para comprobar la independencia entre una variable continua y una variable categórica. Utilice el procedimiento de exploración para comprobar la normalidad de una variable continua. Utilice el procedimiento de prueba de chi-cuadrado para comprobar su una variable categórica tiene especificada una distribución multinomial.

Para obtener un análisis de clústeres en dos fases

1. En los menús seleccione:

   Analizar > **Clasificar** > **Clúster de bietápico...**

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione una o varias variables categóricas o continuas.

Si lo desea, puede:

- Ajustar los criterios utilizados para construir los clústeres.
- Seleccionar los ajustes para el tratamiento del ruido, la asignación de memoria, la estandarización de las variables y la entrada del modelo de clúster.
- Solicitar resultados del visor de modelos.
- Guardar los resultados del modelo en el archivo de trabajo o en un archivo XML externo.

Opciones del análisis de clústeres en dos fases

Tratamiento del valor atípico. Este grupo permite tratar los valores atípicos de manera especial durante la agrupación en clústeres si se llena el árbol de características de los clústeres (CF). El árbol CF se considera lleno si no puede aceptar ningún caso más en un nodo hoja y no hay ningún nodo hoja que se pueda dividir.

- Si selecciona el tratamiento del ruido y el árbol CF se llena, se hará volver a crecer después de colocar los casos existentes en hojas poco densas en una hoja de "ruido". Se considera que una hoja es poco densa si contiene un número de casos inferior a un determinado porcentaje de casos del máximo tamaño de hoja. Tras volver a hacer crecer el árbol, los valores atípicos se colocarán en el árbol CF en caso de que sea posible. Si no es así, se descartarán los valores atípicos.
- Si no selecciona el tratamiento del ruido y el árbol CF se llena, se hará volver a crecer utilizando un umbral del cambio en distancia mayor. Tras la agrupación en clústeres final, los valores que no se puedan asignar a un clúster se considerarán valores atípicos. Al clúster de valores atípicos se le asigna un número de identificación de -1 y no se incluirá en el recuento del número de clústeres.

Asignación de memoria. Este grupo permite especificar la cantidad máxima de memoria en megabytes (MB) que puede utilizar el algoritmo de agrupación en clústeres. Si el procedimiento supera este máximo, utilizará el disco para almacenar la información que no se pueda colocar en la memoria. Especifique un número mayor o igual que 4.

- Consulte con el administrador del sistema si desea conocer el valor máximo que puede especificar en su sistema.
- Si este valor es demasiado bajo, es posible que el algoritmo no consiga obtener el número correcto de clústeres.

Tipificación de variables. El algoritmo de agrupación en clústeres trabaja con variables continuas tipificadas. Todas las variables continuas que no estén tipificadas deben dejarse como variables en la lista Para tipificar. Para ahorrar algún tiempo y trabajo para el ordenador, puede seleccionar todas las variables continuas que ya haya tipificado como variables en la lista Asumidas como tipificadas.

Opciones avanzadas

Criterios de ajuste del árbol CF. Los siguientes ajustes del algoritmo de agrupación en clústeres se aplican específicamente al árbol de características de clústeres (CF) y deberán cambiarse con cuidado:

- **Umbral del cambio en distancia inicial.** Éste es el umbral inicial que se utiliza para hacer crecer el árbol CF. Si se ha insertado una determinada hoja en el árbol CF que produciría una densidad inferior al umbral, la hoja no se dividirá. Si la densidad supera el umbral, se dividirá la hoja.
- **Nº máximo de ramas (por nodo hoja).** Número máximo de nodos hijo que puede tener un nodo hoja.
- **Máxima profundidad de árbol.** Número máximo de niveles que puede tener un árbol CF.
- **Máximo número posible de nodos.** Indica el número máximo de nodos del árbol CF que puede generar potencialmente el procedimiento, de acuerdo con la función $(b^{d+1} - 1) / (b - 1)$, donde b es el número máximo de ramas y d es la profundidad máxima del árbol. Tenga en cuenta que un árbol CF excesivamente grande puede agotar los recursos del sistema y afectar negativamente al rendimiento del procedimiento. Como mínimo, cada nodo requiere 16 bytes.

Actualización del modelo de clúster. Este grupo permite importar y actualizar un modelo de clúster generado en un análisis anterior. El archivo de entrada contiene el árbol CF en formato XML. A continuación, se actualizará el modelo con los datos existentes en el archivo activo. Debe seleccionar los nombres de variable en el cuadro de diálogo principal en el mismo orden en que se especificaron en el análisis anterior. El archivo XML permanecerá inalterado, a no ser que escriba específicamente la nueva información del modelo en el mismo nombre de archivo. Consulte el tema "Resultados de análisis de clústeres en dos fases" en la página 227 para obtener más información.

Si se ha especificado una actualización del modelo de clúster, se utilizarán las opciones pertenecientes a la generación del árbol CF que se especificaron para el modelo original. Concretamente, se utilizarán los ajustes del modelo guardado acerca de la medida de distancia, el tratamiento del ruido, la asignación de memoria y los criterios de ajuste del árbol CF, por lo que se ignorarán todos los ajustes de estas opciones que se hayan especificado en los cuadros de diálogo.

Nota: al realizar una actualización del modelo de clúster, el procedimiento supone que ninguno de los casos seleccionados en el conjunto de datos activo se utilizó para crear el modelo de clúster original. El procedimiento también supone que los casos utilizados en la actualización del modelo proceden de la misma población que los casos utilizados para crear el modelo; es decir, se supone que las medias y las varianzas de las variables continuas y los niveles de las variables categóricas son los mismos en ambos conjuntos de casos. Si los conjuntos de casos "nuevo" y "antiguo" proceden de poblaciones heterogéneas, deberá ejecutar el procedimiento Análisis de clústeres en dos fases para los conjuntos combinados de casos para obtener los resultados óptimos.

Resultados de análisis de clústeres en dos fases

Resultado. Este grupo proporciona opciones para la presentación los resultados de la agrupación en clústeres.

- **Tablas dinámicas.** Los resultados se muestran en tablas dinámicas.
- **Gráficos y tablas en el Visor de modelos.** Los resultados se muestran en el Visor de modelos.
- **Campos de evaluación.** Calcula los datos del clúster de las variables que no se han utilizado en su creación. Los campos de evaluación se pueden mostrar junto con las características de entrada del visor de modelos seleccionándolas en el cuadro de diálogo subordinado Visualización. Los campos con valores perdidos se ignoran.

Archivo de datos de trabajo. Este grupo permite guardar las variables en el conjunto de datos activo.

- **Crear variable del clúster de pertenencia.** Esta variable contiene un número de identificación de clúster para cada caso. El nombre de esta variable es *tsc_n*, donde *n* es un número entero positivo que indica el ordinal de la operación de almacenamiento del conjunto de datos activo realizada por este procedimiento en una determinada sesión.

Archivos XML. El modelo de clúster final y el árbol CF son dos tipos de archivos de resultados que se pueden exportar en formato XML.

- **Exportar modelo final.** También se puede exportar el modelo de clúster final al archivo especificado en formato XML (PMML). Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo.
- **Exportar árbol CF.** Esta opción le permite guardar el estado actual del árbol de clúster y actualizarlo posteriormente utilizando datos más recientes.

El visor de clústeres

Los modelos de clúster se suelen utilizar para buscar grupos (o clústeres) de registros similares basados en las variables examinadas, donde la similitud entre los miembros del mismo grupo es alta y es baja entre miembros de grupos diferentes. Los resultados pueden utilizarse para identificar las asociaciones que, de otra manera, no serían aparentes. Por ejemplo, mediante el análisis de clústeres de preferencias del cliente, de nivel de ingresos y de hábitos de consumo, se podría identificar los tipos de clientes con más probabilidad de responder a una campaña de marketing particular.

Existen dos métodos para interpretar los resultados de una presentación de clústeres:

- Examinar los clústeres para determinar las características exclusivas de cada clúster. *¿Contiene uno de los clústeres todos los socios con un alto nivel de ingresos? ¿Este clúster tiene más registros que los demás?*
- Examinar los campos de todos los clústeres para determinar la forma en que los valores se distribuyen en ellos. *¿Determina el nivel de educación la pertenencia a un clúster? ¿Distingue la puntuación de crédito alto entre la pertenencia a un clúster o a otro?*

Puede utilizar las vistas principales y las diferentes vistas vinculadas en el visor de clústeres para obtener una mayor perspectiva que le ayuda a responder a estas preguntas.

Si desea ver información sobre el modelo de clúster, active el objeto Visor de modelos pulsando dos veces sobre él en el visor.

Visor de clústeres

El Visor de clústeres se compone de dos paneles, la vista principal en la parte izquierda y la vista relacionada o auxiliar de la derecha. Hay dos vistas principales:

- Resumen del modelo (valor predeterminado). Consulte el tema “Vista Resumen del modelo” en la página 228 para obtener más información.
- Clústeres. Consulte el tema “Vista de clústeres” en la página 229 para obtener más información.

Hay cuatro vistas relacionadas/auxiliares:

- Importancia del predictor. Consulte el tema “Vista Importancia del predictor de clústeres” en la página 230 para obtener más información.
- Tamaños de clústeres (valor predeterminado) Consulte el tema “Vista de tamaños de clústeres” en la página 230 para obtener más información.
- Distribución de casillas. Consulte el tema “Vista Distribución de casillas” en la página 231 para obtener más información.
- Comparación de clústeres. Consulte el tema “Vista Comparación de clústeres” en la página 231 para obtener más información.

Vista Resumen del modelo

La vista Resumen del modelo muestra una instantánea o resumen del modelo de clúster, incluyendo una medida de silueta de la cohesión y separación de clústeres sombreada para indicar resultados pobres, correctos o buenos. Esta instantánea le permite comprobar rápidamente si la calidad es insuficiente, en cuyo caso puede optar por volver al nodo de modelado para cambiar los ajustes del modelo de clúster para producir mejores resultados.

Los resultados serán pobres, correctos o buenos de acuerdo con el trabajo de Kaufman y Rousseeuw (1990) sobre la interpretación de estructuras de clústeres. En la vista Resumen del modelo, un resultado "bueno" indica que los datos reflejan una evidencia razonable o sólida de que existe una estructura de clústeres, de acuerdo con la valoración Kaufman y Rousseeuw; una resultado "correcto" indica que ese evidencia es débil, y un resultado "pobre" significa que, según esa valoración, no hay evidencias obvias.

Las medias de medida de silueta, en todos los registros, (B–A) / max(A,B), donde A es la distancia del registro al centro de su clúster y B es la distancia del registro al centro del clúster más cercano al que no pertenece. Un coeficiente de silueta de 1 podría implicar que todos los casos están ubicados directamente en los centros de sus clústeres. Un valor de -1 significaría que todos los casos se encuentran en los centros de clúster de otro clúster. Un valor de 0 implica, de media, que los casos están equidistantes entre el centro de su propio clúster y el siguiente clúster más cercano.

El resumen incluye una tabla que contiene la siguiente información:

- **Algoritmo.** El algoritmo de agrupación en clústeres utilizado, por ejemplo, "Dos fases".
- **Características de entrada.** El número de campos, también conocidos como **entradas** o **predictores**.
- **Clústeres.** Número de clústeres de la solución.

Vista de clústeres

La vista Clústeres contiene una cuadrícula de clústeres por características que incluye nombres de clústeres, tamaños y perfiles para cada clúster.

Las columnas de la cuadrícula contienen la siguiente información:

- **Clúster.** Números de clústeres creados por el algoritmo.
- **Etiqueta.** Etiquetas aplicadas a cada clúster (está en blanco de forma predeterminada). Pulse dos veces la casilla para introducir una etiqueta que describa el contenido del clúster, por ejemplo "Compradores de automóviles de lujo".
- **Descripción.** Cualquier descripción de los contenidos de los clústeres (está en blanco de forma predeterminada). Pulse dos veces la casilla para introducir una descripción del clúster, por ejemplo "Más de 55 años de edad, profesionales, con ingresos superiores a 100.000 €".
- **Tamaño.** El tamaño de cada clúster como porcentaje de la muestra general del clúster. Cada casilla de tamaño de la cuadrícula muestra una barra vertical que muestra el porcentaje de tamaño del clúster, un porcentaje de tamaño en formato numérico y los recuentos de casos de clúster.
- **Características.** Los predictores o entradas individuales, ordenados por importancia general de forma predeterminada. Si hay columnas con tamaños iguales, se muestran en orden ascendente en función de los miembros del clúster.

  La importancia general de la característica se indica por el color del sombreado del fondo de la casilla, siendo más oscuro cuanto más importante sea la característica. Una guía sobre la tabla indica la importancia vinculada a cada color de casilla de característica.

Cuando pasa el ratón por una casilla, se muestra el nombre completo/etiqueta de la característica y el valor de importancia de la casilla. Es posible que aparezca más información, en función de la vista y tipo de característica. En la vista Centros de clústeres, esto incluye la estadística de casilla y el valor de la casilla, por ejemplo: “Media: 4,32”. En las características categóricas, la casilla muestra el nombre de la categoría (modal) más frecuente y su porcentaje.

En la vista Clústeres, puede seleccionar varias formas de mostrar la información de clústeres:

- Transponer clústeres y características Consulte el tema “Transponer clústeres y características” en la página 229 para obtener más información.
- Clasificar características Consulte el tema “Clasificar características” en la página 229 para obtener más información.
- Clasificar clústeres Consulte el tema “Clasificar clústeres” en la página 230 para obtener más información.
- Seleccionar contenido de casilla Consulte el tema “Contenido de casilla” en la página 230 para obtener más información.

Transponer clústeres y características

De forma predeterminada, los clústeres se muestran como columnas y las características se muestran como filas. Para invertir esta visualización, pulse el botón **Transponer clústeres y características** a la izquierda de los botones **Clasificar características**. Por ejemplo, puede que desea hacer esto cuando se muestren muchos clústeres para reducir la cantidad de desplazamiento horizontal necesario para visualizar los datos.

Clasificar características

Los botones **Clasificar características por** le permiten seleccionar la cantidad de casillas de características:

- **Importancia global** Este es el orden de clasificación predeterminado. Las características se clasifican en orden descendente de importancia general y el orden de clasificación es el mismo entre los distintos clústeres. Si hay características que empatan en valores de importancia, éstas se muestran en orden de clasificación ascendente según el nombre.

- **Importancia dentro del clúster** Las características se clasifican con respecto de su importancia para cada clúster. Si hay características que empatan en valores de importancia, éstas se muestran en orden de clasificación ascendente según el nombre. Si esta opción está seleccionada, el orden de clasificación suele variar en los diferentes clústeres.
- **Nombre.** Las características se clasifican por nombre en orden alfabético.
- **Orden de los datos** Las características se clasifican por orden en el conjunto de datos.

Clasificar clústeres

De forma predeterminada, los clústeres se clasifican en orden de tamaño descendente. Los botones **Clasificar clústeres por** le permiten ordenarlos por nombre en orden alfabético o, si ha creado etiquetas de exclusivas, por orden de etiqueta alfanumérico.

Las características con la misma etiqueta se clasifican por nombre de clúster. Si los clústeres se clasifican por etiqueta y modifica la etiqueta de un clúster, el orden de clasificación se actualiza automáticamente.

Contenido de casilla

Los botones **Casillas** le permiten cambiar la visualización del contenido de casillas de características y campos de evaluación.

- **Centros de los clústeres.** De forma predeterminada, las casillas muestran nombres/etiquetas de características y la tendencia central para cada combinación de clúster/característica. La media se muestra para los campos continuos y el modo (categoría más frecuente) con porcentaje de categoría para los campos categóricos.
- **Distribuciones absolutas.** Muestra nombres/etiquetas de características y distribuciones absolutas de las características de cada clúster. En el caso de las características categóricas, la visualización muestra gráficos de barras superpuestas con las categorías ordenadas en orden ascendente de valores de datos. En las características continuas, la visualización muestra un gráfico de densidad suave que utiliza los mismos puntos finales e intervalos para cada clúster.

  La visualización en color rojo oscuro muestra la distribución de clústeres, mientras que la más clara representa los datos generales.

- **Distribuciones relativas** Muestra los nombres/etiquetas de características y las distribuciones relativas en las casillas. En general, las visualizaciones son similares a las mostradas para las distribuciones absolutas, sólo que en su lugar se muestran distribuciones relativas.

  La visualización en color rojo oscuro muestra la distribución de clústeres, mientras que la más clara representa los datos generales.

- **Vista básica.** Si hay muchos clústeres, puede resultar difícil ver todos los detalles sin desplazarse. Para reducir la cantidad de desplazamiento, seleccione esta vista para cambiar la visualización a una versión más compacta de la tabla.

Vista Importancia del predictor de clústeres

La vista Importancia del predictor muestra la importancia relativa de cada campo en la estimación del modelo.

Vista de tamaños de clústeres

La vista Tamaños de clústeres muestra el gráfico circular que contiene cada clúster. El tamaño de porcentaje de cada clúster se muestra en cada sector, pase el ratón sobre cada porción para mostrar el recuento de esa porción.

Bajo el gráfico, una tabla enumera la siguiente información de tamaño:

- El tamaño del clúster más pequeño (un recuento y porcentaje del conjunto).
- El tamaño del clúster mayor (un recuento y porcentaje del conjunto).
- La proporción entre el tamaño del mayor clúster y el del menor.

Vista Distribución de casillas

La vista Distribución de casillas muestra un gráfico expandido y más detallado de la distribución de los datos para cualquier casilla que seleccione en el panel principal Clústeres.

Vista Comparación de clústeres

La vista Comparación de clústeres se compone de un diseño en estilo de cuadrícula, con características en las filas y clústeres seleccionados en las columnas. Esta vista le ayuda a entender mejor los factores de los que se componen los clústeres, y le permite ver las diferencias entre los clústeres no sólo con respecto a los datos generales, sino entre sí.

Para seleccionar clústeres para su visualización, pulse en la parte superior de la columna del clúster en el panel principal Clústeres. Pulse las teclas Ctrl o Mayús y pulse para seleccionar o cancelar la selección de más de un clúster para su comparación.

Nota: Puede seleccionar hasta cinco clústeres para su visualización.

Los clústeres se muestran en el orden en que se seleccionaron, mientras que el orden de los campos viene determinado por la opción **Clasificar características por**. Si selecciona **Importancia dentro del clúster**, los campos siempre se clasifican por importancia general.

Los gráficos de fondo muestran las distribuciones generales de cada característica:

- Las características categóricas aparecen como gráficos de puntos, donde el tamaño del punto indica la categoría más frecuente/modal para cada clúster (por característica).
- Las características continuas se muestran como diagramas de caja, que muestran las medianas globales y los rangos intercuartiles.

En estas vistas de fondo aparecen superpuestos diagramas de caja para los clústeres seleccionados:

- En las características continuas hay marcadores de puntos cuadrados y líneas horizontales que indican el rango de mediana e intercuartil de cada clúster.
- Cada clúster viene representado por un color distinto, que se muestra en la parte superior de la vista.

Navegación en el Visor de clústeres

El visor de clústeres es una pantalla interactiva. Puede:

- Seleccione un campo o clúster para ver más detalles.
- Comparar clústeres para seleccionar elementos de interés.
- Alterar la visualización.
- Transponer ejes.

Uso de las barras de herramientas

Puede controlar la información que aparece en los paneles izquierdo y derecho mediante las opciones de la barra de herramientas. Puede cambiar la orientación de la pantalla (de arriba abajo, de izquierda a derecha, o de derecha a izquierda) mediante los controles de la barra de herramientas. Además, también puede restablecer el visor a los ajustes predeterminados, y abrir un cuadro de diálogo para especificar el contenido de la vista Clústeres en el panel principal.

Las opciones **Clasificar características por**, **Clasificar clústeres por**, **Casillas** y **Mostrar** sólo están disponibles cuando selecciona la vista **Clústeres** en el panel principal. Consulte el tema "Vista de clústeres" en la página 229 para obtener más información.

Tabla 4. Iconos de barra de herramientas

Icono	Tema
	Consulte Transponer clústeres y características

Tabla 4. Iconos de barra de herramientas (continuación)	
Icono	**Tema**
	Consulte Clasificar características por
	Consulte Clasificar clústeres por
	Consulte Casillas

Control de la visualización de clústeres

Para controlar qué se muestra en la vista Clústeres del panel principal, pulse el botón **Mostrar** y se abrirá el cuadro de diálogo Mostrar.

Características. Está seleccionado de forma predeterminada. Para ocultar todas las características de entrada, cancele la selección de la casilla de verificación.

Campos de evaluación Seleccione los campos de evaluación (campos que no se usan para crear el modelo de clúster, sino que se envían al visor de modelos para evaluar los clústeres) que desea mostrar, ya que ninguno se muestra de forma predeterminada. *Nota* El campo de evaluación debe ser una cadena con más de un valor. Este recuadro de selección no está disponible si no hay campos de evaluación disponibles.

Descripciones de clústeres Está seleccionado de forma predeterminada. Para ocultar todas las casillas de descripción de clúster, cancele la selección de la casilla de verificación.

Tamaños de clúster Está seleccionado de forma predeterminada. Para ocultar todas las casillas de tamaño de clúster, cancele la selección de la casilla de verificación.

Número máximo de categorías Especifique el número máximo de categorías que se mostrarán en gráficos de características categóricas. El valor predeterminado es 20.

Filtrado de registros

Si desea obtener más información sobre los casos de un determinado clúster o grupo de clústeres, puede seleccionar un subconjunto de registros para realizar un análisis más detallado en los clústeres seleccionados.

1. Seleccione los clústeres en la vista Clúster del Visor de clústeres. Mantenga pulsada la tecla Ctrl al mismo tiempo que pulsa el botón del ratón para seleccionar varios clústeres.
2. En los menús seleccione:

   Generar > **Filtrar registros...**
3. Introduzca un nombre de variable de filtro. Los registros de los clústeres seleccionados recibirán un valor igual a 1 para este campo. Todos los demás registros recibirán un valor igual a 0 y se excluirán de los análisis subsiguientes hasta que se modifique el estado del filtro.
4. Pulse **Aceptar**.

Análisis de clústeres jerárquico

Este procedimiento intenta identificar grupos relativamente homogéneos de casos (o de variables) basándose en las características seleccionadas, mediante un algoritmo que comienza con cada caso (o cada variable) en un clúster diferente y combina los clústeres hasta que sólo queda uno. Es posible analizar las variables brutas o elegir de entre una variedad de transformaciones de estandarización. Las medidas de distancia o similaridad se generan mediante el procedimiento Proximidades. Los estadísticos se muestran en cada etapa para ayudar a seleccionar la mejor solución.

Ejemplo. ¿Existen grupos identificables de programas televisivos que atraigan a audiencias similares dentro de cada grupo? Con el análisis de clústeres jerárquico, podría agrupar los programas de TV (los casos) en grupos homogéneos basados en las características del espectador. Esto se puede utilizar para

identificar segmentos de mercado. También puede agrupar ciudades (los casos) en grupos homogéneos, de manera que se puedan seleccionar ciudades comparables para probar diversas estrategias de marketing.

Estadísticas. Historial de conglomeración, matriz de distancias (o similaridades) y pertenencia a los clústeres para una solución única o una serie de soluciones. Gráficos: dendrogramas y diagramas de témpanos.

Análisis de clústeres jerárquico: Consideraciones sobre los datos

Datos. Las variables pueden ser cuantitativas, binarias o datos de recuento. El escalamiento de las variables es un aspecto importante, ya que las diferencias en el escalamiento pueden afectar a las soluciones en clústeres. Si las variables muestran grandes diferencias en el escalamiento (por ejemplo, una variable se mide en dólares y la otra se mide en años), debería considerar la posibilidad de estandarizarlas (esto puede llevarse a cabo automáticamente mediante el propio procedimiento Análisis de clústeres jerárquico).

Orden de casos. Si hay distancias empatadas o similitudes en los datos de entrada o si éstas se producen entre los clústeres actualizados durante la unión, la solución de clúster resultante puede depender del orden de los casos del archivo. Puede que desee obtener varias soluciones distintas con los casos ordenados en distintos órdenes aleatorios para comprobar la estabilidad de una solución determinada.

Supuestos. Las medidas de distancia o similaridad empleadas deben ser adecuadas para los datos analizados (véase el procedimiento Proximidades para obtener más información sobre la elección de las medidas de distancia y similaridad). Asimismo, debe incluir todas las variables relevantes en el análisis. Si se omiten variables de interés la solución obtenida puede ser equívoca. Debido a que el análisis de clústeres jerárquico es un método exploratorio, los resultados deben considerarse provisionales hasta que sean confirmados mediante otra muestra independiente.

Para obtener un análisis de clústeres jerárquico

1. En los menús seleccione:

   Analizar > **Clasificar** > **Clústeres jerárquicos...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Si está aglomerando casos, seleccione al menos una variable numérica. Si está aglomerando variables, seleccione al menos tres variables numéricas.

Si lo desea, puede seleccionar una variable de identificación para etiquetar los casos.

Análisis de clústeres jerárquico: Método

Método de agrupación en clústeres. Las opciones disponibles son: Vinculación inter-grupos, Vinculación intra-grupos, Vecino más próximo, Vecino más lejano, Agrupación de centroides, Agrupación de medianas y Método de Ward.

Medida. Permite especificar la medida de distancia o similaridad que será empleada en la aglomeración. Seleccione el tipo de datos y la medida de distancia o similaridad adecuada:

- **Intervalo.** Distancia euclídea, Distancia euclídea al cuadrado, Coseno, Correlación de Pearson, Chebychev, Bloque, Minkowski y Personalizada.
- **Recuentos.** Las opciones disponibles son: Medida de chi-cuadrado y Medida de phi-cuadrado.
- **Binaria.** Las alternativas disponibles son distancia euclídea, distancia euclídea al cuadrado, diferencia de tamaño, diferencia de patrón, varianza, dispersión, forma, coincidencia simple, correlación de 4 puntos de phi, lambda, *D* de Anderberg, dice, Hamann, Jaccard, Kulczynski 1, Kulczynski 2, Lance y Williams, Ochiai, Rogers y Tanimoto, Russel y Rao, Sokal y Sneath 1, Sokal y Sneath 2, Sokal y Sneath 3, Sokal y Sneath 4, Sokal y Sneath 5, *Y* de Yule y *Q* de Yule.

Transformar valores. Permite estandarizar los valores de los datos, para los casos o las variables, antes de calcular las proximidades (no está disponible para datos binarios). Los métodos disponibles de

estandarización son: Puntuaciones *z*, Rango -1 a 1, Rango 0 a 1, Magnitud máxima de 1, Media de 1 y Desviación estándar 1.

Transformar medidas. Permite transformar los valores generados por la medida de distancia. Se aplican después de calcular la medida de distancia. Las opciones disponibles son: Valores absolutos, Cambiar el signo y Cambiar la escala al rango 0–1.

Análisis de clústeres jerárquico: Estadísticos

Historial de conglomeración. Muestra los casos o clústeres combinados en cada etapa, las distancias entre los casos o los clústeres que se combinan, así como el último nivel del proceso de aglomeración en el que cada caso (o variable) se unió a su clúster correspondiente.

Matriz de proximidades. Proporciona las distancias o similaridades entre los elementos.

Clúster de pertenencia. Muestra el clúster al cual se asigna cada caso en una o varias etapas de la combinación de los clústeres. Las opciones disponibles son: Solución única y Rango de soluciones.

Análisis de clústeres jerárquico: Gráficos

Dendrograma. Muestra un *dendrograma*. Los dendrogramas pueden emplearse para evaluar la cohesión de los clústeres que se han formado y proporcionar información sobre el número adecuado de clústeres que deben conservarse.

Témpanos. Muestra un *diagrama de témpanos*, que incluye todos los clústeres o un rango especificado de clústeres. Los diagramas de témpanos muestran información sobre cómo se combinan los casos en los clústeres, en cada iteración del análisis. La orientación permite seleccionar un diagrama vertical u horizontal.

Análisis de clústeres jerárquico: Guardar variables nuevas

Clúster de pertenencia. Permite guardar los clústeres de pertenencia para una solución única o un rango de soluciones. Las variables guardadas pueden emplearse en análisis posteriores para explorar otras diferencias entre los grupos.

Características adicionales de la sintaxis de comandos CLUSTER

El procedimiento Clúster jerárquico utiliza la sintaxis de comandos CLUSTER. La sintaxis de comandos también le permite:

- Utilizar varios métodos de agrupación en un único análisis.
- Leer y analizar una matriz de proximidades.
- Escribir una matriz de distancias para su análisis posterior.
- Especificar cualquier valor para la potencia y la raíz en la medida de distancia personalizada (potencia).
- Especificar nombres para variables guardadas.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Análisis de clústeres de K-medias

Este procedimiento intenta identificar grupos de casos relativamente homogéneos basándose en las características seleccionadas y utilizando un algoritmo que puede gestionar un gran número de casos. Sin embargo, el algoritmo requiere que el usuario especifique el número de clústeres. Puede especificar los centros iniciales de los clústeres si conoce de antemano dicha información. Puede elegir uno de los dos métodos disponibles para clasificar los casos: la actualización de los centros de los clústeres de forma iterativa o sólo la clasificación. Asimismo, puede guardar la pertenencia a los clústeres, información de la distancia y los centros de los clústeres finales. Si lo desea, puede especificar una variable cuyos valores sean utilizados para etiquetar los resultados por casos. También puede solicitar los estadísticos *F* de los análisis de varianza. Aunque estos estadísticos son oportunistas (ya que el procedimiento trata de formar

grupos que de hecho difieran), el tamaño relativo de los estadísticos proporciona información acerca de la contribución de cada variable a la separación de los grupos.

Ejemplo. ¿Cuáles son los grupos identificables de programas de televisión que atraen audiencias parecidas dentro de cada grupo? Con el análisis de clústeres de *k*-medias, podría agrupar los programas de televisión (los casos) en *k* grupos homogéneos, basados en las características del televidente. Este proceso se puede utilizar para identificar segmentos de mercado. También puede agrupar ciudades (los casos) en grupos homogéneos, de manera que se puedan seleccionar ciudades comparables para probar diversas estrategias de marketing.

Estadísticas. Solución completa: centros iniciales de los clústeres, tabla de ANOVA. Cada caso: información del clúster, distancia desde el centro del clúster.

Análisis de clústeres de K-medias: Consideraciones sobre los datos

Datos. Las variables deben ser cuantitativas en el nivel de intervalo o de razón. Si las variables son binarias o recuentos, utilice el procedimiento Análisis de clústeres jerárquicos.

Orden de casos y centro de clústeres iniciales. El algoritmo predeterminado para elegir centros de clústeres iniciales no es invariable con respecto a la ordenación de casos. La opción **Usar medias actualizadas** del cuadro de diálogo Iterar hace que la solución resultante dependa potencialmente del orden de casos con independencia de cómo se eligen los centros de clústeres iniciales. Si va a utilizar alguno de estos métodos, puede que desee obtener varias soluciones distintas con los casos ordenados en distintos órdenes aleatorios para comprobar la estabilidad de una solución determinada. La especificación de los centros de clústeres iniciales y la no utilización de la opción **Usar medias actualizadas** evita los problemas relacionados con el orden de casos. No obstante, la ordenación de los centros de clústeres iniciales puede afectar a la solución en caso de haber distancias empatadas desde los casos a los centros de clústeres. Para evaluar la estabilidad de una solución determinada, puede comparar los resultados de los análisis con las distintas permutaciones de los valores de centros iniciales.

Supuestos. Las distancias se calculan utilizando la distancia euclídea simple. Si desea utilizar otra medida de distancia o de similaridad, utilice el procedimiento Análisis de clústeres jerárquicos. El escalamiento de variables es una consideración importante. Si sus variables utilizan diferentes escalas (por ejemplo, una variable se expresa en dólares y otra, en años), los resultados podrían ser equívocos. En estos casos, debería considerar la estandarización de las variables antes de realizar el análisis de clústeres de *k*-medias (esta tarea se puede hacer en el procedimiento Descriptivos). Este procedimiento supone que ha seleccionado el número apropiado de clústeres y que ha incluido todas las variables relevantes. Si ha seleccionado un número inapropiado de clústeres o ha omitido variables relevantes, los resultados podrían ser equívocos.

Para obtener un análisis de clústeres de K-medias

1. En los menús seleccione:

   Analizar > **Clasificar** > **Clúster de K medias...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione las variables que se van a utilizar en el análisis de clústeres.
3. Especifique el número de clústeres. (Este número no debe ser inferior a 2 ni superior al número de casos del archivo de datos.)
4. Seleccione **Iterar y clasificar** o **Clasificar sólo**.
5. Si lo desea, seleccione una variable de identificación para etiquetar los casos.

Eficacia del análisis de clústeres de K-medias

El comando de análisis de clústeres de *k*-medias es eficaz principalmente porque no calcula las distancias entre todos los pares de casos, como hacen muchos algoritmos de agrupación en clústeres, como el utilizado por el comando de agrupación en clústeres jerárquica.

Para conseguir la máxima eficacia, tome una muestra de los casos y seleccione el método **Iterar y clasificar** para determinar los centros de los clústeres. Seleccione **Escribir finales en**. A continuación,

restaure el archivo de datos completo, seleccione el método **Sólo clasificar** y seleccione **Leer iniciales de** para clasificar el archivo completo utilizando los centros estimados a partir de la muestra. Se puede escribir y leer desde un archivo o conjunto de datos. Los conjuntos de datos están disponibles para su uso posterior durante la misma sesión, pero no se guardarán como archivos a menos que se hayan guardado explícitamente antes de que finalice la sesión. El nombre de un conjunto de datos debe cumplir las normas de denominación de variables. Para obtener más información, consulte el tema .

Análisis de clústeres de K-medias: Iterar

Nota: estas opciones sólo están disponibles si se selecciona el método **Iterar y clasificar** en el cuadro de diálogo Análisis de clústeres de K-medias.

Número máximo de iteraciones. Limita el número de iteraciones en el algoritmo *k*-medias. La iteración se detiene después de este número de iteraciones, incluso si no se ha satisfecho el criterio de convergencia. Este número debe estar entre el 1 y el 999.

Para reproducir el algoritmo utilizado por el comando Quick Cluster en las versiones previas a la 5.0, establezca **Máximo de iteraciones** en 1.

Criterio de convergencia. Determina cuándo cesa la iteración. Representa una proporción de la distancia mínima entre los centros de clúster iniciales, por lo que debe ser mayor que 0 pero no mayor que 1. Si el criterio es igual a 0.02, por ejemplo, la iteración cesa cuando una iteración completa no mueve ninguno de los centros de clúster por una distancia de más del 2 % de la distancia más pequeña entre los centros de clúster iniciales.

Usar medias actualizadas. Permite solicitar la actualización de los centros de los clústeres tras la asignación de cada caso. Si no selecciona esta opción, los nuevos centros de los clústeres se calcularán después de la asignación de todos los casos.

Análisis de clústeres de K-medias: Guardar

Puede guardar información sobre la solución como nuevas variables para que puedan ser utilizadas en análisis subsiguientes:

Clúster de pertenencia. Crea una nueva variable que indica el clúster final al que pertenece cada caso. Los valores de la nueva variable van desde el 1 hasta el número de clústeres.

Distancia desde centro del clúster. Crea una nueva variable que indica la distancia euclídea entre cada caso y su centro de clasificación.

Análisis de clústeres de K-medias: Opciones

Estadísticas. Puede seleccionar los siguientes estadísticos: Centros de clústeres iniciales, Tabla de ANOVA e Información del clúster para cada caso.

- *Centros de clústeres iniciales*. Primera estimación de las medias variables para cada uno de los clústeres. De forma predeterminada, se selecciona entre los datos un número de casos debidamente espaciados igual al número de clústeres. Los centros iniciales de los clústeres se utilizan como criterio para una primera clasificación y, a partir de ahí, se van actualizando.
- *tabla de ANOVA*. Muestra una tabla de análisis de varianza que incluye pruebas F univariadas para cada variable de agrupación en clúster. Las pruebas F son sólo descriptivas y las probabilidades resultantes no se deben interpretar. La tabla de ANOVA no se mostrará si se asignan todos los casos a un único clúster.
- *Información de clúster para cada caso*. Muestra para cada caso la asignación de clúster final y la distancia euclídea entre el caso y el centro de clúster utilizado para clasificar el caso. También muestra la distancia euclídea entre los centros de los clústeres finales.

Valores perdidos. Las opciones disponibles son: **Excluir casos según lista** o **Excluir casos según pareja**.

- **Excluir casos según lista.** Excluye los casos con valores perdidos para cualquier variable de agrupación del análisis.

- **Excluir casos según pareja.** Asigna casos a los clústeres en función de las distancias que se calculan desde todas las variables con valores no perdidos.

Características adicionales del comando QUICK CLUSTER

El procedimiento K-Means Cluster utiliza la sintaxis de comandos `QUICK CLUSTER` . La sintaxis de comandos también le permite:

- Aceptar los primeros casos *k* como primeros centros de clústeres iniciales y, por lo tanto, evitar la lectura de datos que normalmente se utiliza para calcularlos.
- Especificar los centros de clústeres iniciales directamente como parte de la sintaxis de comandos.
- Especificar nombres para variables guardadas.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Pruebas no paramétricas

Las pruebas no paramétricas hacen supuestos mínimos acerca de la distribución subyacente de los datos. Las pruebas que están disponibles en estos cuadros de diálogo, se pueden agrupar en tres categorías amplias en función de cómo se organizan los datos:

- Una prueba de una muestra analiza un campo.
- Una prueba para muestras relacionadas compara dos o más campos para el mismo conjunto de casos.
- Una prueba de muestras independientes analiza un campo que se agrupa por categorías de otro campo.

Pruebas no paramétricas de una muestra

Una prueba no paramétricas de una muestra identifica diferencias en campos únicos mediante una o más pruebas no paramétricas. Las pruebas no paramétricas no dan por hecho que sus datos sigan la distribución normal.

¿Cuál es su objetivo? Los objetivos le permiten especificar rápidamente ajustes de prueba diferentes y comunes.

- **Comparar automáticamente datos observados con el valor hipotetizado.** Este objetivo aplica la prueba binomial a campos categóricos con sólo dos categorías, la prueba de chi-cuadrado al resto de campos categóricos y la prueba de Kolmogorov-Smirnov a campos continuos.
- **Probar la aleatoriedad de la secuencia.** Este objetivo utiliza la prueba de rachas para comprobar la aleatoriedad de la secuencia observada de valores de datos.
- **Análisis personalizado.** Seleccione esta opción si desea modificar manualmente la configuración de la prueba de la pestaña Configuración. Tenga en cuenta que esta configuración se selecciona automáticamente si realiza cambios posteriores a muchas opciones de la pestaña Configuración que sean incompatibles con los del objetivo seleccionado actualmente.

Obtención de pruebas no paramétricas para una muestra

En los menús seleccione:

Analizar > **Pruebas no paramétricas** > **Una muestra...**

1. Pulse **Ejecutar**.

Si lo desea, puede:

- Especifique un objetivo en la pestaña Objetivos.
- Especifique asignaciones de campo en la pestaña Campos.
- Especifique los valores de experto en la pestaña Valores.

Pestaña Campos

La pestaña Campos especifica los campos que se deben comprobar.

Utilizar papeles predefinidos. Esta opción utiliza información de campos existentes. Todos los campos con un papel predefinido como Entrada o Ambos se utilizarán como campos de prueba. Al menos un campo de prueba es necesario.

Utilizar asignaciones de campos personalizadas. Esta opción le permite sobrescribir papeles de campos. Después de seleccionar esta opción, especifique los campos siguientes.

- **Campos de prueba.** Seleccione uno o más campos de prueba.

Pestaña Configuración

La pestaña Configuración contiene diferentes grupos de ajustes que puede modificar para ajustar con precisión la forma en que el algoritmo procesa sus datos. Si realiza algún cambio en la configuración predeterminada que sea incompatible con el objetivo seleccionado actualmente, la pestaña Objetivo se actualiza automáticamente para seleccionar la opción **Personalizar análisis**.

Seleccionar pruebas

Estos ajustes especifican las pruebas que realizarán en los campos especificados en la pestaña Campos.

Seleccione automáticamente las pruebas en función de los datos. Esta configuración aplica la prueba binomial a campos categóricos con sólo dos categorías válidas (sin valores perdidos), la prueba de chi-cuadrado al resto de campos categóricos y la prueba de Kolmogorov-Smirnov a campos continuos.

Personalizar pruebas. Esta configuración permite especificar las pruebas que se ejecutarán.

- **Comparar la probabilidad binaria observada con el valor hipotetizado (prueba binomial).** La prueba binomial se puede aplicar a todos los campos. Produce una prueba de una muestra que comprueba si la distribución observada de un campo de distintivo (un campo categórico con sólo dos categorías) es el mismo que lo que se espera de una distribución binomial especificada. Además, puede solicitar intervalos de confianza. Consulte “Opciones de prueba binomial ” en la página 238 para obtener más información sobre la configuración de prueba.
- **Comparar las probabilidades observadas con el valor hipotetizado (prueba de chi-cuadrado).** La prueba de chi-cuadrado se aplica a campos nominales y ordinales. Produce una prueba de una muestra que calcula un estadístico chi-cuadrado basado en las diferencias entre las frecuencias observadas y esperadas de las categorías de un campo. Consulte “Opciones de prueba de chi-cuadrado ” en la página 239 para obtener más información sobre la configuración de prueba.
- **Probar la distribución observada con el valor hipotetizado (prueba de Kolmogorov-Smirnov).** La prueba de Kolmogorov-Smirnov se aplica a campos continuos y ordinales. Produce una prueba de una muestra de si la función de distribución acumulada de muestra de un campo es homogénea con una distribución uniforme, normal, Poisson o exponencial. Consulte “Opciones de Kolmogorov-Smirnov ” en la página 239 para obtener más información sobre la configuración de prueba.
- **Comparar mediana con el valor hipotetizado (prueba de Wilcoxon de los rangos con signo).** La prueba de Wilcoxon de los rangos con signo se aplica a los campos continuos y ordinales. Produce una prueba para una muestra del valor de mediana de un campo. Especifique un número como la mediana hipotetizada.
- **Probar la aleatoriedad de la secuencia (prueba de rachas).** La prueba de rachas se aplica a todos los campos. Produce una prueba de una muestra de si la secuencia de valores de un campo de dicotomías es aleatoria. Consulte “Prueba de rachas: Opciones” en la página 240 para obtener más información sobre la configuración de prueba.

Opciones de prueba binomial

La prueba binomial está diseñada para campos de distintivo (campos categóricos con sólo dos categorías), pero se aplica a todos los campos mediante reglas para definir "éxito".

Proporción hipotetizada. Especifica la proporción esperada de registros definidos como "éxitos" o *p*. Especifique un valor mayor que 0 y menor que 1. El valor predeterminado es 0,5.

Intervalo de confianza. Los siguientes métodos permiten calcular intervalos de confianza de datos binarios:

- **Clopper-Pearson (exacto).** Un intervalo exacto basado en la distribución binomial acumulada.
- **Jeffreys.** Un intervalo Bayesian basado en la distribución posterior de *p* que utiliza la opción Jeffreys anterior.
- **Razón de verosimilitud.** Un intervalo basado en la función de verosimilitud para *p*.

Definir éxito para campos categóricos. Especifica cómo se define "éxito", el valor(s) de datos se comprueba en la proporción hipotetizada, en los campos categóricos.

- **Utilizar primera categoría encontrada en los datos** realiza la prueba binomial que utiliza el primer valor encontrado en la muestra para definir "éxito". Esta opción sólo es aplicable a los campos nominal u ordinal con sólo dos valores; el resto de campos categóricos especificados en la pestaña Campos en los que se utiliza esta opción no se comprobarán. Este es el valor predeterminado.
- **Especificar valores de éxito** realiza la prueba binomial que utiliza la lista especificada de valores para definir "éxito". Especifique una lista de valores de cadena o numérico. No es necesario que los valores de la lista estén en la muestra.

Definir éxito para campos continuos. Especifica cómo se define "éxito", el valor(s) de datos se comprueba en el valor de prueba, en los campos categóricos. Éxito se define como valores iguales o menores que un punto de corte.

- **Punto medio de muestra** define el punto de corte en la media de los valores mínimo o máximo.
- **Punto de corte personalizado** permite especificar un valor para el punto de corte.

Opciones de prueba de chi-cuadrado

Todas las categorías tienen la misma probabilidad. Produce la misma frecuencia entre todas las categorías en la muestra. Este es el valor predeterminado.

Personalizar probabilidad esperada. Permite especificar frecuencias desiguales para una lista de categorías específica. Especifique una lista de valores de cadena o numérico. No es necesario que los valores de la lista estén en la muestra. En la columna **Categoría**, especifique los valores de categorías. En la columna **Frecuencia relativa**, especifique un valor superior a 0 para cada categoría. Las frecuencias personalizadas se consideran porcentajes de forma que, por ejemplo, especificar frecuencias de 1, 2 y 3 es equivalente a especificar frecuencias de 10, 20 y 30, y especificar que 1/6 de los registros se esperan en la primera categoría, 1/3 en la segunda y 1/2 en la tercera. Si se especifican probabilidades esperadas personalizadas, los valores de categorías personalizadas deben incluir todos los valores de campo de los datos; de lo contrario la prueba no se realiza en ese campo.

Opciones de Kolmogorov-Smirnov

Este cuadro de diálogo especifica las distribuciones que se deben comprobar y los parámetros de las distribución hipotetizada.

Cuando deben estimarse algunos parámetros de la distribución a partir de la muestra, la prueba de Kolmogorov-Smirnov ya no se aplica. En estas instancias, la estadística de prueba de Lilliefors se puede utilizar para estimar el valor *p* utilizando el muestreo de Monte Carlo para contrastar la normalidad con la media y la varianza desconocida. La prueba de Lilliefors se aplica a las tres distribuciones continuas (**Normal**, **Exponencial** y **Uniforme**). Tenga en cuenta que la prueba no se aplica si la distribución subyacente es discreta (**Poisson**). La prueba solo está definida para inferencia de una muestra cuando no se especifican los parámetros de distribución correspondientes.

Normal

Utilizar datos de muestra utiliza la media observada y la desviación estándar y proporciona opciones para seleccionar los resultados de **Prueba asintótica** o utilice **Prueba de Lilliefors basado en muestreo de Monte Carlo**. **Personalizado** le permite especificar valores.

Uniforme

Utilizar datos de muestra utiliza los valores mínimos y máximos observados y utiliza la prueba de and uses Lilliefors basada en el muestreo de Monte Carlo. **Personalizado** le permite especificar valores mínimos y máximos.

Exponencial

Media muestral utiliza la media observada y utiliza la prueba de Lilliefors basada en el muestreo de Monte Carlo. **Personalizado** le permite especificar un valor medio observado.

Poisson

Media le permite especificar un valor medio observado.

Prueba de rachas: Opciones

La prueba de rachas está diseñada para campos de distintivo (campos categóricos con sólo dos categorías), pero se puede aplica a todos los campos mediante reglas para definir grupos.

Definir grupos para campos categóricos. Se encuentran disponibles las siguientes opciones:

- **Sólo hay 2 categorías en la muestra** realiza la prueba de rachas utilizando los valores encontrados en la muestra para definir los grupos. Esta opción sólo es aplicable a los campos nominal u ordinal con sólo dos valores; el resto de campos categóricos especificados en la pestaña Campos en los que se utiliza esta opción no se comprobarán.
- **Recodificar datos en 2 categorías** realiza la prueba de rachas utilizando la lista de valores especificada para definir uno de los grupos. El resto de valores de muestra definen el otro grupo. No es necesario que todos los valores de la lista estén presentes en la muestra, pero debe haber al menos un registro en cada grupo.

Definir punto de corte para campos continuos. Especifica cómo se definen los grupos para campos continuos. El primer grupo se define como valores iguales o menores que un punto de corte.

- **Mediana muestral** define el punto de corte en la mediana muestral.
- **Media muestral** define el punto de corte en la media muestral.
- **Personalizado** permite especificar un valor para el punto de corte.

Opciones de prueba

Nivel de significación

Especifica el nivel de significación (alfa) de todas las pruebas. Especifique un valor numérico entre 0 y 1. 0,05 es el valor predeterminado.

Intervalo de confianza (%)

Esto especifica el nivel de confianza de todos los intervalos de confianza producidos. Especifique un valor numérico entre 0 y 100. 95 es el valor predeterminado.

Casos excluidos

Especifica cómo se determinan las pruebas caso por caso.

Excluir casos según prueba

Los registros con valores perdidos para un campo que se utiliza para una prueba específica se omiten de esta prueba. Si se realizan varias pruebas en el análisis, cada prueba se evalúa por separado.

Excluir casos según lista

Los registros con valores perdidos para cualquier campo que se nombre en la pestaña Campos se excluyen de todos los análisis.

Muestreo de Monte Carlo

Cuando deben estimarse algunos parámetros de la distribución a partir de la muestra, la prueba de Kolmogorov-Smirnov ya no se aplica. En estas instancias, la estadística de prueba de Lilliefors se puede utilizar para estimar el valor *p* utilizando el muestreo de Monte Carlo para contrastar la normalidad con la media y la varianza desconocida. La prueba de Lilliefors se aplica a las tres distribuciones continuas (**Normal**, **Exponencial** y **Uniforme**). Tenga en cuenta que la prueba no se

aplica si la distribución subyacente es discreta (**Poisson**). La prueba solo está definida para inferencia de una muestra cuando no se especifican los parámetros de distribución correspondientes.

Establecer semilla personalizada
Cuando está habilitado, este valor proporciona la opción de restablecer el valor de **Semilla** que se utiliza para el muestreo de Monte Carlo. El valor debe ser un entero único entre 1 y 2,147,483,647 (el valor predeterminado es 2,000,000). De forma predeterminada, el valor no está habilitado, lo que produce un valor de inicio aleatorio.

Número de muestras
Restablece el número de réplicas de muestreo de Monte Carlo que se utilizan en la prueba de Lilliefors. El valor debe ser un único entero entre 100 y el entero más grande. El valor predeterminado es 10.000.

Nivel de confianza de simulación (%)
Restablece el nivel de intervalo de confianza estimado de la prueba de Kolmogorov-Smirnov. El valor debe estar comprendido por un único valor entre 0 y 100. El valor predeterminado es 99.

Valores perdidos del usuario

Valores perdidos del usuario para campos categóricos. Para que un registro se incluya en el análisis, los campos categóricos deben tener valores válidos para dicho caso. Estos controles permiten decidir si los valores perdidos del usuario se deben tratar como válidos entre los campos categóricos. Los valores perdidos del sistema y los valores perdidos de campos continuos siempre se tratan como no válidos.

Características adicionales del comando NPTESTS

La sintaxis de comandos también le permite:

- Especifique pruebas de una muestra, muestras independientes y muestras relacionadas en una única ejecución del procedimiento.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Prueba de Wilcoxon de una muestra

La prueba de Wilcoxon de los rangos con signo es una prueba estadística no paramétrica que se utiliza para probar la ubicación de una población basándose en los datos de la muestra o para comparar las ubicaciones de dos poblaciones. La prueba de Wilcoxon de los rangos con signo de una muestra es una alternativa no paramétrica a la prueba t de la muestra cuando los datos no siguen una distribución normal. Ayuda a determinar si existe una diferencia significativa entre la mediana de una muestra y un valor de mediana hipotetizado. Aquí, el estadístico de prueba se calcula basándose en los rangos de la diferencia entre los valores observados y el valor hipotetizado.

La prueba de Wilcoxon de los rangos con signo de una muestra da por supuesto que la población padre es simétrica respecto a la mediana. Como las muestras aleatorias provienen de una población con una función de distribución acumulada continua y son simétricas sobre la mediana, bajo la hipótesis nula, las diferencias entre los valores observados y el valor hipotetizado se distribuyen simétricamente alrededor de cero de manera que las diferencias positivas y negativas de igual magnitud absoluta tienen la misma probabilidad de ocurrencia.

Ejemplo

Supongamos que una nueva empresa está interesada en contratar ingenieros de software, pero desea asegurarse de que están ofreciendo salarios competitivos. Supongamos que han recopilado datos salariales de una muestra de 30 ingenieros de software que actualmente están empleados en varias empresas de la misma industria. Si la empresa desea determinar si el salario medio que están ofreciendo a sus potenciales contrataciones es significativamente diferente del salario medio de los ingenieros de software en la industria, entonces se puede utilizar un rango de signo de Wilcoxon de muestra para este propósito. Aquí, asumimos la hipótesis nula de que el salario mediano ofrecido por la startup no es significativamente diferente del salario mediano en la industria, es decir, la diferencia salarial mediana es cero.

Para ejecutar la prueba de Wilcoxon de una muestra

(Esta característica requiere la opción Statistics Base.)

En el menú, elija:

1. **Analizar > Pruebas no paramétricas > Una muestra ...**
2. Pulse **'Configuración'**.
3. En **'Seleccionar un elemento'**, seleccione **'Seleccionar pruebas'**.
4. Seleccione **'Personalizar pruebas'**.
5. Seleccione **'Comparar mediana con hipotetizado (prueba de rango con signo Wilcoxon)'**.
6. Especifique el valor de **'Mediana hipotetizada'** y pulse **'Ejecutar'**.

Para obtener más información, consulte https://www.ibm.com/support/pages/can-spss-statistics-produce-one-sample-wilcoxon-test

Pruebas no paramétricas de muestras independientes

Las pruebas no paramétricas de muestras independientes identifican diferencias entre dos o más grupos utilizando una o más pruebas no paramétricas. Las pruebas no paramétricas no dan por hecho que sus datos sigan la distribución normal.

¿Cuál es su objetivo? Los objetivos le permiten especificar rápidamente ajustes de prueba diferentes y comunes.

- **Comparar automáticamente distribuciones entre grupos.** Este objetivo aplica la prueba U de Mann-Whitney para datos con 2 grupos o la prueba ANOVA de 1 factor de Kruskal-Wallis para datos con grupos *k* .
- **Comparar medianas entre grupos.** Este objetivo utiliza la prueba de la mediana para comparar las medianas observadas entre grupos.
- **Análisis personalizado.** Seleccione esta opción si desea modificar manualmente la configuración de la prueba de la pestaña Configuración. Tenga en cuenta que esta configuración se selecciona automáticamente si realiza cambios posteriores a muchas opciones de la pestaña Configuración que sean incompatibles con los del objetivo seleccionado actualmente.

Para obtener pruebas no paramétricas de muestras independientes

En los menús seleccione:

Analizar > **Pruebas no paramétricas** > **Muestras independientes...**

1. Pulse **Ejecutar**.

Si lo desea, puede:

- Especifique un objetivo en la pestaña Objetivos.
- Especifique asignaciones de campo en la pestaña Campos.
- Especifique la configuración de experto en la pestaña Configuración.

Pestaña Campos

La pestaña Campos especifica los campos que se deben comprobar y el campo que se utilizará para definir grupos.

Utilizar papeles predefinidos. Esta opción utiliza información de campos existentes. Todos los campos continuos y ordinales con un rol predefinido como Destino o Ambos se utilizarán como campos de prueba. Si hay un único campo categórico con un papel predefinido como Entrada se utilizará como un campo de agrupación. De lo contrario, no se utilizará de forma predeterminada ningún campo de agrupación y deberá utilizar asignaciones de campos personalizadas. Se requiere al menos un campo de prueba y un campo de agrupación.

Utilizar asignaciones de campos personalizadas. Esta opción le permite sobrescribir papeles de campos. Después de seleccionar esta opción, especifique los campos siguientes.

- **Campos de prueba.** Seleccione uno o más campos continuos u ordinales.
- **Grupos.** Seleccione un campo categórico.

Pestaña Configuración

La pestaña Configuración contiene diferentes grupos de ajustes que puede modificar para ajustar con precisión con la que el algoritmo procesa sus datos. Si realiza algún cambio en la configuración predeterminada que sea incompatible con el objetivo seleccionado actualmente, la pestaña Objetivo se actualiza automáticamente para seleccionar la opción **Personalizar análisis**.

Seleccionar pruebas

Estos ajustes especifican las pruebas que realizarán en los campos especificados en la pestaña Campos.

Seleccione automáticamente las pruebas en función de los datos. Esta configuración aplica la prueba U de Mann-Whitney para datos con 2 grupos o la prueba ANOVA de 1 factor de Kruskal-Wallis para datos con grupos *k* .

Personalizar pruebas. Esta configuración permite especificar las pruebas que se ejecutarán.

- **Comparar distribuciones entre grupos.** Producen pruebas para muestras independientes si las muestras son de la misma población.

  U de Mann-Whitney (2 muestras) utiliza el nivel de cada caso para comprobar si los grupos se extraen de la misma población. El primer valor del campo de agrupación define el grupo de control y el segundo define el grupo de comparación. Si el campo de agrupación tiene más de dos valores, esta prueba no se ejecuta.

  Kolmogorov-Smirnov (2 muestras) es sensible a cualquier diferencia en la mediana, dispersión, asimetría, etcétera entre las dos distribuciones. Si el campo de agrupación tiene más de dos valores, esta prueba no se ejecuta.

  Probar la aleatoriedad de la secuencia (Wald-Wolfowitz para 2 muestras) produce una prueba de rachas con la pertenencia al grupo como criterio. Si el campo de agrupación tiene más de dos valores, esta prueba no se ejecuta.

  ANOVA de 1 factor de Kruskal-Wallis (k muestras) es una extensión de la prueba U de Mann-Whitney y el análogo no paramétrico de análisis de varianza de un factor. Opcionalmente puede solicitar múltiples comparaciones de las muestras *k*, en comparaciones múltiples **todo por parejas** o comparaciones **por pasos en sentido descendente**.

  Probar alternativas ordenadas (prueba de Jonckheere-Terpstra para k muestras) es una alternativa más potente que Kruskal-Wallis si las muestras *k* tienen un orden natural. Por ejemplo, las *k* poblaciones pueden representar *k* temperaturas ascendentes. Se contrasta la hipótesis de que diferentes temperaturas producen la misma distribución de respuesta, con la hipótesis alternativa de que cuando la temperatura aumenta, la magnitud de la respuesta aumenta. La hipótesis alternativa se encuentra aquí ordenada; por tanto, la prueba de Jonckheere-Terpstra es la prueba más apropiada. **Del menor al mayor** especifica la hipótesis alternativa de que el parámetro de ubicación del primer grupo es menor o igual que el segundo, que es menor o igual que el tercero, etc. **Del mayor al menor** especifica la hipótesis alternativa de que el parámetro de ubicación del primer grupo es mayor o igual que el segundo, que es mayor o igual que el tercero, etc. Para ambas opciones, la hipótesis alternativa también asume que las ubicaciones no son todas iguales. Opcionalmente puede solicitar múltiples comparaciones de las muestras *k*, en comparaciones múltiples **todo por parejas** o comparaciones **por pasos en sentido descendente**.
- **Comparar rangos entre grupos.** Produce una prueba de muestras independientes de si las muestras tienen el mismo rango. **Reacciones extremas de Moses (2 muestras)** comprueba un grupo de control con un grupo de comparación. El primer valor en orden ascendente del campo de agrupación define el grupo de control y el segundo define el grupo de comparación. Si el campo de agrupación tiene más de dos valores, esta prueba no se ejecuta.

- **Comparar medianas entre grupos.** Produce una prueba de muestras independientes de si las muestras tienen la misma mediana. **Prueba de la mediana (k muestras)** puede utilizar la mediana muestral combinada (calculada con todos los registros del conjunto de datos) o un valor personalizado como la mediana hipotetizada. Opcionalmente puede solicitar múltiples comparaciones de las muestras *k*, en comparaciones múltiples **todo por parejas** o comparaciones **por pasos en sentido descendente**.
- **Estimar intervalos de confianza entre grupos. Hodges-Lehman (2 muestras)** produce una estimación muestras independientes y el intervalo de confianza para la diferencia en las medianas de los dos grupos. Si el campo de agrupación tiene más de dos valores, esta prueba no se ejecuta.

Opciones de prueba

Nivel de significación. Especifica el nivel de significación (alfa) de todas las pruebas. Especifique un valor numérico entre 0 y 1. 0,05 es el valor predeterminado.

Intervalo de confianza (%). Especifica el nivel de confianza para todos los intervalos de confianza generados. Especifique un valor numérico entre 0 y 100. 95 es el valor predeterminado.

Casos excluidos. Especifica cómo se determinan las pruebas caso por caso. **Excluir casos según lista** significa que los registros con valores perdidos de cualquier campo que se nombran en cualquier subcomando se excluyen de todos los análisis. **Excluir casos según prueba** significa que los registros con valores perdidos para un campo que se utiliza para una prueba específica se omiten de esa prueba. Si se realizan varias pruebas en el análisis, cada prueba se evalúa por separado.

Valores perdidos del usuario

Valores perdidos del usuario para campos categóricos. Para que un registro se incluya en el análisis, los campos categóricos deben tener valores válidos para dicho caso. Estos controles permiten decidir si los valores perdidos del usuario se deben tratar como válidos entre los campos categóricos. Los valores perdidos del sistema y los valores perdidos de campos continuos siempre se tratan como no válidos.

Características adicionales del comando NPTESTS

La sintaxis de comandos también le permite:

- Especifique pruebas de una muestra, muestras independientes y muestras relacionadas en una única ejecución del procedimiento.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Pruebas no paramétricas de muestras relacionadas

Identifica diferencias entre dos o más campos relacionados mediante una o más pruebas no paramétricas. Las pruebas no paramétricas no dan por hecho que sus datos sigan la distribución normal.

Consideraciones sobre los datos. Cada registro corresponde a un sujeto concreto para el que se almacenan dos o más mediciones relacionadas en campos separados del conjunto de datos. Por ejemplo, es posible analizar un estudio sobre la efectividad de un plan de dietas mediante pruebas no paramétricas de muestras relacionadas si el peso de cada sujeto se mide a intervalos regulares y se almacena como campos como *Peso previo a la dieta*, *Peso intermedio* y *Peso tras la dieta*. Estos campos están "relacionados".

¿Cuál es su objetivo? Los objetivos le permiten especificar rápidamente ajustes de prueba diferentes y comunes.

- **Comparar automáticamente datos observados con datos hipotetizados.** Este objetivo aplica la prueba de McNemar a datos categóricos cuando se especifican 2 campos, la prueba Q de Cochran datos categóricos cuando se especifican más de 2 campos, la prueba de Wilcoxon de los rangos con signo a datos continuos cuando se especifican 2 campos y ANOVA de 2 vías de Friedman por rangos a datos continuos cuando se especifican más de 2 campos.
- **Análisis personalizado.** Seleccione esta opción si desea modificar manualmente la configuración de la prueba de la pestaña Configuración. Tenga en cuenta que esta configuración se selecciona

automáticamente si realiza cambios posteriores a muchas opciones de la pestaña Configuración que sean incompatibles con los del objetivo seleccionado actualmente.

Cuando se especifican campos de diferentes niveles de medición, primero se separan por nivel de medición y después se aplica la prueba adecuada a cada grupo. Por ejemplo, si selecciona **Comparar automáticamente datos observados con el valor hipotetizado** como objetivo y especifica 3 campos continuos y 2 campos nominales, se aplicará la prueba de Friedman a los campos continuos y la prueba de McNemar a los campos nominales.

Para obtener pruebas no paramétricas para muestras relacionadas

En los menús seleccione:

Analizar > **Pruebas no paramétricas** > **Muestras relacionadas...**

1. Pulse **Ejecutar**.

Si lo desea, puede:

- Especifique un objetivo en la pestaña Objetivos.
- Especifique asignaciones de campo en la pestaña Campos.
- Especifique la configuración de experto en la pestaña Configuración.

Pestaña Campos

La pestaña Campos especifica los campos que se deben comprobar.

Utilizar papeles predefinidos. Esta opción utiliza información de campos existentes. Todos los campos con un papel predefinido como Destino o Ambos se utilizarán como campos de prueba. Se requieren al menos dos campos de prueba.

Utilizar asignaciones de campos personalizadas. Esta opción le permite sobrescribir papeles de campos. Después de seleccionar esta opción, especifique los campos siguientes.

- **Campos de prueba.** Seleccione dos o más campos. Cada campo corresponde a una muestra relacionada diferente.

Pestaña Configuración

La pestaña Configuración contiene diferentes grupos de ajustes que puede modificar para ajustar con precisión con la que el procedimiento procesa sus datos. Si realiza algún cambio en la configuración predeterminada que sea incompatible con el resto de objetivos, la pestaña Objetivo se actualiza automáticamente para seleccionar la opción**Personalizar análisis**.

Seleccionar pruebas

Estos ajustes especifican las pruebas que realizarán en los campos especificados en la pestaña Campos.

Seleccione automáticamente las pruebas en función de los datos. Esta configuración aplica la prueba de McNemar a datos categóricos cuando se especifican 2 campos, la prueba Q de Cochran datos categóricos cuando se especifican más de 2 campos, la prueba de Wilcoxon de los rangos con signo a datos continuos cuando se especifican 2 campos y ANOVA de 2 vías de Friedman por rangos a datos continuos cuando se especifican más de 2 campos.

Personalizar pruebas. Esta configuración permite especificar las pruebas que se ejecutarán.

- **Prueba de cambio en datos binarios** La **prueba de McNemar (2 muestras)** se puede aplicar a campos categóricos. Produce una prueba de muestras relacionadas de si las combinaciones de valores entre dos campos de distintivo (campos categóricos con dos valores únicamente) son igualmente probables. Si hay más de dos campos especificados en la pestaña Campos, esta prueba no se realiza. Consulte “Prueba de McNemar: definir éxito ” en la página 246 para obtener más información sobre la configuración de prueba. **Q de Cochran (k muestras)** se puede aplicar a campos categóricos. Produce una prueba de muestras relacionadas de si las combinaciones de valores entre *k* campos de distintivo

(campos categóricos con dos valores únicamente) son igualmente probables. Opcionalmente puede solicitar múltiples comparaciones de las muestras *k*, en comparaciones múltiples **todo por parejas** o comparaciones **por pasos en sentido descendente**. Consulte “Prueba de Cochran: definir éxito ” en la página 246 para obtener más información sobre la configuración de prueba.

- **Prueba de cambios en datos multinomiales** La **prueba de homogeneidad marginal (2 muestras)** produce una prueba de muestras relacionadas de si las combinaciones de valores entre dos campos ordinales emparejados son igualmente probables. La prueba de homogeneidad marginal se suele utilizar en situaciones de medidas repetidas. Se trata de una extensión de la prueba de McNemar a partir de la respuesta binaria a la respuesta multinomial. Si hay más de dos campos especificados en la pestaña Campos, esta prueba no se realiza.
- **Comparar diferencia de la mediana con el valor hipotetizado.** Estas pruebas producen cada una una prueba de muestras relacionadas sobre si la diferencia mediana entre dos es diferente de 0. Esta prueba se aplica a campos continuos y ordinales. Si hay más de dos campos especificados en la pestaña Campos, estas pruebas no se realizan.
- **Estimar intervalo de confianza.** Produce un cálculo de muestras relacionadas y un intervalo de confianza para la diferencia de la mediana entre dos campos emparejados. Esta prueba se aplica a campos continuos y ordinales. Si hay más de dos campos especificados en la pestaña Campos, esta prueba no se realiza.
- **Cuantificar asociaciones. Coeficiente de concordancia de Kendall (k muestras)** produce un coeficiente de concordancia entre evaluadores, donde cada registro es un valor de evaluador de varios elementos (campos). Opcionalmente puede solicitar múltiples comparaciones de las muestras *k*, en comparaciones múltiples **todo por parejas** o comparaciones **por pasos en sentido descendente**.
- **Compare distribuciones ANOVA bidireccional por rangos de Friedman (muestras k)** produce una prueba de muestras relacionadas de si se han extraído muestras relacionadas con *k* de la misma población. Opcionalmente puede solicitar múltiples comparaciones de las muestras *k*, en comparaciones múltiples **todo por parejas** o comparaciones **por pasos en sentido descendente**.

Prueba de McNemar: definir éxito

La prueba de McNemar está diseñada para campos de distintivo (campos categóricos con sólo dos categorías), pero se aplica a todos los campos categóricos mediante reglas para definir "éxito".

Definir éxito para campos categóricos. Especifica cómo se define "éxito" en los campos categóricos.

- **Utilizar primera categoría encontrada en los datos** realiza la prueba que utiliza el primer valor encontrado en la muestra para definir "éxito". Esta opción sólo es aplicable a los campos nominal u ordinal con sólo dos valores; el resto de campos categóricos especificados en la pestaña Campos en los que se utiliza esta opción no se comprobarán. Este es el valor predeterminado.
- **Especificar valores de éxito** realiza la prueba que utiliza la lista especificada de valores para definir "éxito". Especifique una lista de valores de cadena o numérico. No es necesario que los valores de la lista estén en la muestra.

Prueba de Cochran: definir éxito

La prueba de Q de Cochran está diseñada para campos de distintivo (campos categóricos con sólo dos categorías), pero se aplica a todos los campos categóricos mediante reglas para definir "éxito".

Definir éxito para campos categóricos. Especifica cómo se define "éxito" en los campos categóricos.

- **Utilizar primera categoría encontrada en los datos** realiza la prueba que utiliza el primer valor encontrado en la muestra para definir "éxito". Esta opción sólo es aplicable a los campos nominal u ordinal con sólo dos valores; el resto de campos categóricos especificados en la pestaña Campos en los que se utiliza esta opción no se comprobarán. Este es el valor predeterminado.
- **Especificar valores de éxito** realiza la prueba que utiliza la lista especificada de valores para definir "éxito". Especifique una lista de valores de cadena o numérico. No es necesario que los valores de la lista estén en la muestra.

Opciones de prueba

Nivel de significación. Especifica el nivel de significación (alfa) de todas las pruebas. Especifique un valor numérico entre 0 y 1. 0,05 es el valor predeterminado.

Intervalo de confianza (%). Especifica el nivel de confianza para todos los intervalos de confianza generados. Especifique un valor numérico entre 0 y 100. 95 es el valor predeterminado.

Casos excluidos. Especifica cómo se determinan las pruebas caso por caso.

- **Excluir casos según lista** significa que los registros con valores perdidos de cualquier campo que se nombran en cualquier subcomando se excluyen de todos los análisis.
- **Excluir casos según prueba** significa que los registros con valores perdidos para un campo que se utiliza para una prueba específica se omiten de esa prueba. Si se realizan varias pruebas en el análisis, cada prueba se evalúa por separado.

Valores perdidos del usuario

Valores perdidos del usuario para campos categóricos. Para que un registro se incluya en el análisis, los campos categóricos deben tener valores válidos para dicho caso. Estos controles permiten decidir si los valores perdidos del usuario se deben tratar como válidos entre los campos categóricos. Los valores perdidos del sistema y los valores perdidos de campos continuos siempre se tratan como no válidos.

Características adicionales del comando NPTESTS

La sintaxis de comandos también le permite:

- Especifique pruebas de una muestra, muestras independientes y muestras relacionadas en una única ejecución del procedimiento.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Características adicionales del comando NPTESTS

La sintaxis de comandos también le permite:

- Especifique pruebas de una muestra, muestras independientes y muestras relacionadas en una única ejecución del procedimiento.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Cuadros de diálogo antiguos

Hay diferentes cuadros de diálogo "antiguos" que también realizan pruebas no paramétricas. Estos cuadros de diálogo admiten la funcionalidad que ofrece la opción Pruebas exactas.

Prueba de chi-cuadrado. Tabula una variable en categorías y calcula un estadístico de chi-cuadrado basándose en las diferencias entre las frecuencias observadas y las esperadas.

Prueba binomial. Compara la frecuencia observada en cada categoría de una variable dicotómica con las frecuencias esperadas en la distribución binomial.

Prueba de rachas. Comprueba si el orden de aparición de dos valores de una variable es aleatorio.

Prueba de Kolmogorov-Smirnov para una muestra. Compara la función de distribución acumulada observada de una variable con una distribución teórica especificada, que puede ser normal, uniforme, exponencial o de Poisson.

Pruebas para dos muestras independientes. Compara dos grupos de casos en una variable. Se encuentran disponibles la prueba *U* de Mann-Whitney, la prueba de Kolmogorov-Smirnov para dos muestras, la prueba de Moses de reacciones extremas y la prueba de rachas de Wald-Wolfowitz.

Pruebas para dos muestras relacionadas. Compara las distribuciones de dos variables. Se encuentran disponibles la prueba de Wilcoxon de los rangos con signo, la prueba de signos y la prueba de McNemar.

Pruebas para varias muestras independientes. Compara dos o más grupos de casos en una variable. Se encuentran disponibles la prueba de Kruskal-Wallis, la prueba de la mediana y la prueba de Jonckheere-Terpstra.

Pruebas para varias muestras relacionadas. Compara las distribuciones de dos o más variables. Se encuentran disponibles la prueba de Friedman, la *W* de Kendall y la *Q* de Cochran.

Además, para todas las pruebas citadas anteriormente están disponibles los cuartiles y la media, la desviación estándar, el mínimo, el máximo y, por último, el número de casos no perdidos.

Prueba de chi-cuadrado

El procedimiento Prueba de chi-cuadrado tabula una variable en categorías y calcula un estadístico de chi-cuadrado. Esta prueba de bondad de ajuste compara las frecuencias observadas y esperadas en cada categoría para contrastar que todas las categorías contengan la misma proporción de valores o que cada categoría contenga una proporción de valores especificada por el usuario.

Ejemplos. La prueba de chi-cuadrado podría utilizarse para determinar si una bolsa de caramelos contiene en igualdad de proporción caramelos de color azul, marrón, verde, naranja, rojo y amarillo. También podría utilizarse para ver si una bolsa de caramelos contiene un 5% de color azul, un 30% de color marrón, un 10% de color verde, un 20% de color naranja, un 15% de color rojo y un 15% de color amarillo.

Estadísticas. Media, desviación estándar, mínimo, máximo y cuartiles. Número y porcentaje de casos perdidos y no perdidos; número de casos observados y esperados de cada categoría; residuos y estadístico de chi-cuadrado.

Prueba de chi-cuadrado: Consideraciones sobre los datos

Datos. Use variables categóricas numéricas ordenadas o no ordenadas (niveles de medición ordinal o nominal). Para convertir las variables de cadena en variables numéricas, utilice el procedimiento Recodificación automática, disponible en el menú Transformar.

Supuestos. Las pruebas no paramétricas no requieren supuestos sobre la forma de la distribución subyacente. Se asume que los datos son una muestra aleatoria. Las frecuencias esperadas para cada categoría deben ser al menos 1. No más del 20 % de las categorías deberían tener frecuencias esperadas inferiores a 5.

Para obtener una prueba de chi-cuadrado

1. En los menús seleccione:

   Analizar > **Pruebas no paramétricas** > **Cuadros de diálogo antiguos** > **Chi-cuadrado...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables de contraste. Cada variable genera una prueba independiente.
3. Si lo desea, puede pulsar en **Opciones** para obtener estadísticos descriptivos, cuartiles y controlar el tratamiento de los datos perdidos.

Prueba de chi-cuadrado: Rango y valores esperados

Rango esperado. De forma predeterminada, cada valor distinto de la variable se define como una categoría. Para establecer categorías dentro de un rango específico, seleccione **Usar rango especificado** e introduzca valores enteros para los límites inferior y superior. Se establecerán categorías para cada valor entero dentro del rango inclusivo y los casos con valores fuera de los límites se excluirán. Por ejemplo, si se especifica 1 como límite inferior y 4 como límite superior, únicamente se utilizarán los valores enteros entre 1 y 4 para la prueba de chi-cuadrado.

Valores esperados. De forma predeterminada, todas las categorías tienen valores esperados iguales. Las categorías pueden tener proporciones esperadas especificadas por el usuario. Seleccione **Valores**, introduzca un valor mayor que 0 para cada categoría de la variable de contraste y, a continuación, pulse en **Añadir**. Cada vez que se añade un valor, éste aparece al final de la lista de valores. El orden de los valores es importante; corresponde al orden ascendente de los valores de categoría de la variable de

contraste. El primer valor de la lista corresponde al valor de grupo mínimo de la variable de contraste y el último valor corresponde al valor máximo. Los elementos de la lista de valores se suman y, a continuación, cada valor se divide por esta suma para calcular la proporción de casos esperados en la categoría correspondiente. Por ejemplo, una lista de valores de 3, 4, 5, 4 especifica unas proporciones esperadas de 3/16, 4/16, 5/16 y 4/16.

Prueba de chi-cuadrado: Opciones

Estadísticos. Puede elegir uno o los dos estadísticos de resumen.

- **Descriptivos.** Muestra la media, la desviación estándar, el mínimo, el máximo y el número de casos no perdidos.
- **Cuartiles.** Muestra los valores correspondientes a los percentiles 25, 50 y 75.

Valores perdidos. Controla el tratamiento de los valores perdidos.

- **Excluir casos según prueba.** Cuando se especifican varias pruebas, cada una se evalúa separadamente respecto a los valores perdidos.
- **Excluir casos según lista.** Los casos con valores perdidos para cualquier variable se excluyen de todos los análisis.

Características adicionales del comando NPAR TESTS (Prueba de chi-cuadrado)

La sintaxis de comandos también le permite:

- Especificar valores mínimos y máximos o frecuencias esperadas diferentes para diferentes variables (con el subcomando CHISQUARE).
- Contrastar la misma variable respecto a diferentes frecuencias esperadas o utilizar diferentes rangos (con el subcomando EXPECTED).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Prueba binomial

El procedimiento Prueba binomial compara las frecuencias observadas de las dos categorías de una variable dicotómica con las frecuencias esperadas en una distribución binomial con un parámetro de probabilidad especificado. De forma predeterminada, el parámetro de probabilidad para ambos grupos es 0,5. Para cambiar las probabilidades, puede introducirse una proporción de prueba para el primer grupo. La probabilidad del segundo grupo será 1 menos la probabilidad especificada para el primer grupo.

Ejemplo. Cuando lanza una moneda, la probabilidad de que salga cara equivale a 1/2. A partir de esta hipótesis, se lanza una moneda 40 veces, y se registran los resultados (cara o cruz). De la prueba binomial, podría deducir que en 3/4 de los lanzamientos salió cara y que el nivel de significación observado es pequeño (0,0027). Estos resultados indican que no es verosímil que la probabilidad de que salga cara sea 1/2; probablemente la moneda presenta una tendencia a caer por un sentido determinado.

Estadísticas. Media, desviación estándar, mínimo, máximo, número de casos no perdidos y cuartiles.

Prueba binomial: Consideraciones sobre los datos

Datos. Las variables de contraste deben ser numéricas y dicotómicas. Para convertir las variables de cadena en variables numéricas, utilice el procedimiento Recodificación automática, disponible en el menú Transformar. Una **variable dicotómica** es una variable que sólo puede tomar dos valores posibles: *sí* o *no*, *verdadero* o *falso*, 0 o 1, etc. El primer valor encontrado en los datos define el primer grupo y el otro valor define el segundo grupo. Si las variables no son dicotómicas, debe especificar un punto de corte. El punto de corte asigna los casos con valores menores o iguales que el punto de corte del primer grupo y asigna el resto de los casos a un segundo grupo.

Supuestos. Las pruebas no paramétricas no requieren supuestos sobre la forma de la distribución subyacente. Se asume que los datos son una muestra aleatoria.

Para obtener una prueba binomial

1. En los menús seleccione:

Analizar > **Pruebas no paramétricas** > **Cuadros de diálogo antiguos** > **Binomial...**

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione una o más variables de contraste numéricas.
3. Si lo desea, puede pulsar en **Opciones** para obtener estadísticos descriptivos, cuartiles y controlar el tratamiento de los datos perdidos.

Prueba binomial: Opciones

Estadísticos. Puede elegir uno o los dos estadísticos de resumen.

- **Descriptivos.** Muestra la media, la desviación estándar, el mínimo, el máximo y el número de casos no perdidos.
- **Cuartiles.** Muestra los valores correspondientes a los percentiles 25, 50 y 75.

Valores perdidos. Controla el tratamiento de los valores perdidos.

- **Excluir casos según prueba.** Cuando se especifican varias pruebas, cada una se evalúa separadamente respecto a los valores perdidos.
- **Excluir casos según lista.** Se excluirán de todos los análisis los casos con valores perdidos de cualquier variable.

Características adicionales del comando NPAR TESTS (Prueba binomial)

La sintaxis de comandos también le permite:

- Seleccionar grupos específicos (y excluir otros grupos) si una variable tiene más de dos categorías (mediante el subcomando `BINOMIAL`).
- Especificar diferentes probabilidades o puntos de corte para diferentes variables (mediante el subcomando `BINOMIAL`).
- Contrastar la misma variable respecto a diferentes probabilidades o puntos de corte (mediante el subcomando `EXPECTED`).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Prueba de rachas

El procedimiento Prueba de rachas contrasta si es aleatorio el orden de aparición de dos valores de una variable. Una racha es una secuencia de observaciones similares. Una muestra con un número excesivamente grande o excesivamente pequeño de rachas sugiere que la muestra no es aleatoria.

Ejemplos. Suponga que se realiza una encuesta a 20 personas para saber si comprarían un producto. Si todas estas 20 personas fueran del mismo sexo, se pondría seriamente en duda la supuesta aleatoriedad de la muestra. La prueba de rachas se puede utilizar para determinar si la muestra fue extraída de manera aleatoria.

Estadísticas. Media, desviación estándar, mínimo, máximo, número de casos no perdidos y cuartiles.

Prueba de rachas: Consideraciones sobre los datos

Datos. Las variables deben ser numéricas. Para convertir las variables de cadena en variables numéricas, utilice el procedimiento Recodificación automática, disponible en el menú Transformar.

Supuestos. Las pruebas no paramétricas no requieren supuestos sobre la forma de la distribución subyacente. Utilice muestras de distribuciones de probabilidad continua.

Para obtener una prueba de rachas

1. En los menús seleccione:

Analizar > **Pruebas no paramétricas** > **Cuadros de diálogo antiguos** > **Rachas...**

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione una o más variables de contraste numéricas.
3. Si lo desea, puede pulsar en **Opciones** para obtener estadísticos descriptivos, cuartiles y controlar el tratamiento de los datos perdidos.

Prueba de rachas: punto de corte

Punto de corte. Especifica un punto de corte para dicotomizar las variables seleccionadas. Puede utilizar como punto de corte los valores observados para la media, la mediana o el modo, o bien un valor especificado. Los casos con valores menores que el punto de corte se asignarán a un grupo y los casos con valores mayores o iguales que el punto de corte se asignarán a otro grupo. Se lleva a cabo una prueba para cada punto de corte seleccionado.

Prueba de rachas: Opciones

Estadísticos. Puede elegir uno o los dos estadísticos de resumen.

- **Descriptivos.** Muestra la media, la desviación estándar, el mínimo, el máximo y el número de casos no perdidos.
- **Cuartiles.** Muestra los valores correspondientes a los percentiles 25, 50 y 75.

Valores perdidos. Controla el tratamiento de los valores perdidos.

- **Excluir casos según prueba.** Cuando se especifican varias pruebas, cada una se evalúa separadamente respecto a los valores perdidos.
- **Excluir casos según lista.** Los casos con valores perdidos para cualquier variable se excluyen de todos los análisis.

Características adicionales del comando NPAR TESTS (Prueba de rachas)

La sintaxis de comandos también le permite:

- Especificar puntos de corte diferentes para las distintas variables (con el subcomando RUNS).
- Contrastar la misma variable con distintos puntos de corte personalizados (con el subcomando RUNS).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Prueba de Kolmogorov-Smirnov para una muestra

El procedimiento Prueba de Kolmogorov-Smirnov para una muestra compara la función de distribución acumulada observada de una variable con una distribución teórica determinada, que puede ser la normal, la uniforme, la de Poisson o la exponencial. La *Z* de Kolmogorov-Smirnov se calcula a partir de la diferencia mayor (en valor absoluto) entre las funciones de distribución acumuladas teórica y observada. Esta prueba de bondad de ajuste contrasta si las observaciones podrían razonablemente proceder de la distribución especificada.

A partir de la versión 27.0, la estadística de prueba Lilliefors se puede utilizar para estimar el valor *p* utilizando el muestreo de Monte Carlo para probar en una distribución normal con parámetros estimados (esta funcionalidad era posible anteriormente únicamente a través del procedimiento Explorar).

Ejemplo
Muchas pruebas paramétricas requieren que las variables se distribuyan de forma normal. La prueba de Kolmogorov-Smirnov para una muestra se puede utilizar para comprobar que una variable (por ejemplo *ingresos*) se distribuye normalmente.

Estadísticos
Media, desviación estándar, mínimo, máximo, número de casos no perdidos, cuartiles, prueba de Lilliefors y simulación de Monte Carlo.

Consideraciones sobre los datos de prueba de Kolmogorov-Smirnov para una muestra

Datos

Utilice variables cuantitativas (a nivel de medición de razón o de intervalo).

Supuestos

La prueba de Kolmogorov-Smirnov asume que los parámetros de la distribución de prueba se han especificado previamente. Este procedimiento estima los parámetros a partir de la muestra. La media y la desviación estándar de la muestra son los parámetros de una distribución normal, los valores mínimo y máximo de la muestra definen el rango de la distribución uniforme, la media muestral es el parámetro de la distribución de Poisson y la media muestral es el parámetro de la distribución exponencial. La capacidad de la prueba para detectar desviaciones a partir de la distribución hipotetizada puede disminuir gravemente.

Cuando deben estimarse algunos parámetros de la distribución a partir de la muestra, la prueba de Kolmogorov-Smirnov ya no se aplica. En estas instancias, la estadística de prueba de Lilliefors se puede utilizar para estimar el valor *p* utilizando el muestreo de Monte Carlo para contrastar la normalidad con la media y la varianza desconocida. La prueba de Lilliefors se aplica a las tres distribuciones continuas (**Normal**, **Exponencial** y **Uniforme**). Tenga en cuenta que la prueba no se aplica si la distribución subyacente es discreta (**Poisson**). La prueba solo está definida para inferencia de una muestra cuando no se especifican los parámetros de distribución correspondientes.

Obtención de una prueba de Kolmogorov-Smirnov para una muestra

1. En los menús seleccione:

   Analizar > **Pruebas no paramétricas** > **Cuadros de diálogo antiguos** > **K-S de 1 muestra...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables de contraste numéricas. Cada variable genera una prueba independiente.
3. Opcionalmente, seleccione un método de distribución de prueba:

   Normal

   Cuando se selecciona, puede especificar si los parámetros de distribución se calculan a partir de los datos de muestra (el valor predeterminado) o de los valores personalizados. Cuando se selecciona **Utilizar datos de muestra**, se utilizan tanto los resultados asintóticos existentes como la corrección de significación de Lilliefors basada en el muestreo de Monte Carlo. Cuando se selecciona **Personalizado**, proporcione valores para **Media** y **Desv. estándar**.

   Uniforme

   Cuando se selecciona, puede especificar si los parámetros de distribución se calculan a partir de los datos de muestra (el valor predeterminado) o de los valores personalizados. Cuando se selecciona **Utilizar datos de ejemplo**, se utiliza la prueba de Lilliefors. Cuando se selecciona **Personalizado**, proporcione valores tanto para **Mín** como para **Máx**.

   Poisson

   Cuando se selecciona, especifique un valor de parámetro **Media**.

   Exponencial

   Cuando se selecciona, puede especificar si los parámetros de distribución se calculan a partir de la media muestral (el valor predeterminado) o de los valores personalizados. Cuando se selecciona **Utilizar datos de ejemplo**, se utiliza la prueba de Lilliefors. Cuando se selecciona **Personalizado**, especifique un valor de parámetro **Media**.
4. Opcionalmente, pulse **Simulación** para especificar parámetros de simulación de Monte Carlo, pulse **Exacto** para especificar parámetros de prueba exactos y pulse **Opciones** para obtener estadísticas descriptivas, cuartiles y control del tratamiento de los datos perdidos.

Prueba de Kolmogorov-Smirnov de una muestra: Simulación

Cuando deben estimarse algunos parámetros de la distribución a partir de la muestra, la prueba de Kolmogorov-Smirnov ya no se aplica. En estas instancias, la estadística de prueba de Lilliefors se puede utilizar para estimar el valor *p* utilizando el muestreo de Monte Carlo para contrastar la normalidad con la media y la varianza desconocida. La prueba de Lilliefors se aplica a las tres distribuciones continuas (**Normal**, **Exponencial** y **Uniforme**). Tenga en cuenta que la prueba no se aplica si la distribución subyacente es discreta (**Poisson**). La prueba solo está definida para inferencia de una muestra cuando no se especifican los parámetros de distribución correspondientes.

Parámetros de simulación de Monte Carlo

Nivel de confianza
Este valor opcional restablece el nivel de intervalo de confianza que estima la prueba de Kolmogorov-Smirnov cuando se utilizan las simulaciones de Monte Carlo. El valor debe estar comprendido entre 0 y 100. El valor predeterminado es 99.

Número de muestras
Este valor opcional restablece el número de réplicas que utiliza la prueba de Lilliefors para el muestreo de Monte Carlo. El valor debe estar comprendido entre 10000 y el número mayor de valores de la muestra. El valor predeterminado es 10000.

Suprimir los resultados de Monte Carlo para la distribución normal
Este ajuste opcional suprime el muestreo de Monte Carlo para los resultados de distribución normal. De forma predeterminada, el valor no está seleccionado (lo que significa que se presentan los resultados asintóticos existentes y los resultados de prueba de Lilliefors, que se basan en el muestreo de Monte Carlo).

Prueba de Kolmogorov-Smirnov para una muestra: Opciones

Estadísticos
Puede elegir uno o los dos estadísticos de resumen.

Descriptivos
Muestra la media, la desviación estándar, el mínimo, el máximo y el número de casos no perdidos.

Cuartiles
Muestra los valores correspondientes a los percentiles 25, 50 y 75.

Valores perdidos
Controla el tratamiento de los valores perdidos.

Excluir casos según prueba
Cuando se especifican varias pruebas, cada una se evalúa separadamente respecto a los valores perdidos.

Excluir casos según lista
Los casos con valores perdidos para cualquier variable se excluyen de todos los análisis.

Características adicionales del comando NPAR TESTS (Prueba de Kolmogorov-Smirnov para una muestra)

El lenguaje de sintaxis de comandos también permite especificar los parámetros de la distribución de prueba (con el subcomando K-S).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Pruebas para dos muestras independientes

El procedimiento Pruebas para dos muestras independientes compara dos grupos de casos existentes en una variable.

Ejemplo. Se han desarrollado nuevos correctores dentales diseñados para que sean más cómodos y estéticos, así como para facilitar un progreso más rápido en la realineación de la dentadura. Para averiguar si el nuevo corrector debe llevarse tanto tiempo como el modelo antiguo, se eligen 10 niños

al azar para que lleven este último y otros 10 niños para que usen el nuevo. Mediante la prueba *U* de Mann-Whitney podría descubrir que, de media, los niños que llevaban el nuevo corrector tenían que llevarlo puesto menos tiempo que los que llevaban el antiguo.

Estadísticas. Media, desviación estándar, mínimo, máximo, número de casos no perdidos y cuartiles. Pruebas: *U* de Mann-Whitney, reacciones extremas de Moses, *Z* de Kolmogorov-Smirnov, rachas de Wald-Wolfowitz.

Pruebas para dos muestras independientes: Consideraciones sobre los datos

Datos. Utilice variables numéricas que puedan ser ordenables.

Supuestos. Utilice muestras independientes y aleatorias. La prueba Mann-Whitney *U* comprueba la igualdad de las dos distribuciones. Para utilizarla para comprobar sus diferencias en la ubicación entre dos distribuciones, se debe asumir que las distribuciones tienen la misma forma.

Para obtener pruebas para dos muestras independientes

1. En los menús seleccione:

   Analizar > **Pruebas no paramétricas** > **Cuadros de diálogo antiguos** > **2 muestras independientes...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables numéricas.
3. Seleccione una variable de agrupación y pulse en **Definir grupos** para segmentar el archivo en dos grupos o muestras.

Tipos de pruebas para dos muestras independientes.

Tipo de prueba. Hay cuatro pruebas disponibles para contrastar si dos muestras (grupos) independientes proceden de una misma población.

La **prueba U de Mann-Whitney** es la más conocida de las pruebas para dos muestras independientes. Es equivalente a la prueba de la suma de rangos de Wilcoxon y a la prueba de Kruskal-Wallis para dos grupos. La prueba de Mann-Whitney contrasta si dos poblaciones muestreadas son equivalentes en su posición. Las observaciones de ambos grupos se combinan y clasifican, asignándose el rango de promedio en caso de producirse empates. El número de empates debe ser pequeño en relación con el número total de observaciones. Si la posición de las poblaciones es idéntica, los rangos deberían mezclarse aleatoriamente entre las dos muestras. La prueba calcula el número de veces que una puntuación del grupo 1 precede a una puntuación del grupo 2 y el número de veces que una puntuación del grupo 2 precede a una puntuación del grupo 1. El estadístico *U* de Mann-Whitney es el más pequeño de estos dos números. También se muestra el estadístico *W* de la suma de rangos de Wilcoxon. *W* es la suma de las rachas del grupo con el rango medio más pequeño, a menos que los grupos tengan el mismo rango promedio, en cuyo caso es la suma de rango del grupo que se denomina último en el cuadro de diálogo Grupos de definición de dos muestras independientes.

La **prueba Z de Kolmogorov-Smirnov** y la **prueba de rachas de Wald-Wolfowitz** son pruebas más generales que detectan las diferencias entre las posiciones y las formas de las distribuciones. La prueba de Kolmogorov-Smirnov se basa en la diferencia máxima absoluta entre las funciones de distribución acumulada observadas para ambas muestras. Cuando esta diferencia es significativamente grande, se consideran diferentes las dos distribuciones. La prueba de rachas de Wald-Wolfowitz combina y ordena las observaciones de ambos grupos. Si las dos muestras son de la misma población, los dos grupos deben estar dispersos al azar a lo largo de la clasificación.

La **prueba de reacciones extremas de Moses** presupone que la variable experimental afectará a algunos sujetos en una dirección y a otros sujetos en la dirección opuesta. La prueba contrasta las respuestas extremas comparándolas con un grupo de control. Esta prueba se centra en la amplitud del grupo de control y supone una medida de la influencia de los valores extremos del grupo experimental en la amplitud al combinarse con el grupo de control. El grupo de control se define en el cuadro Grupo 1 del cuadro de diálogo Dos muestras independientes: Definir grupos. Las observaciones de ambos grupos se combinan y ordenan. El intervalo del grupo de control se calcula como la diferencia entre los rangos de los valores más grandes y más pequeños del grupo de control más 1. Dado que los valores

atípicos pueden distorsionar fácilmente el rango del intervalo, el 5 % de los casos de control se recorta automáticamente de cada extremo.

Pruebas para dos muestras independientes: Definir grupos

Para dividir el archivo en dos grupos o muestras, especifique un valor entero para el Grupo 1 y otro valor para el Grupo 2. Los casos con otros valores se excluyen del análisis.

Pruebas para dos muestras independientes: Opciones

Estadísticos. Puede elegir uno o los dos estadísticos de resumen.

- **Descriptivos.** Muestra la media, la desviación estándar, el mínimo, el máximo y el número de casos no perdidos.
- **Cuartiles.** Muestra los valores correspondientes a los percentiles 25, 50 y 75.

Valores perdidos. Controla el tratamiento de los valores perdidos.

- **Excluir casos según prueba.** Cuando se especifican varias pruebas, cada una se evalúa separadamente respecto a los valores perdidos.
- **Excluir casos según lista.** Los casos con valores perdidos para cualquier variable se excluyen de todos los análisis.

Características adicionales del comando NPAR TESTS (Dos muestras independientes)

El lenguaje de sintaxis de comandos permite especificar el número de casos que se recortarán en la prueba de Moses (con el subcomando MOSES).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Pruebas para dos muestras relacionadas

El procedimiento Pruebas para dos muestras relacionadas compara las distribuciones de dos variables.

Ejemplo. En general, cuando una familia vende su casa ¿logra obtener la cantidad que pide inicialmente? Si aplica la prueba de Wilcoxon de los rangos con signo a 10 casas, podría descubrir que siete familias reciben menos cantidad de la solicitada, una recibe más y dos familias reciben el precio solicitado.

Estadísticas. Media, desviación estándar, mínimo, máximo, número de casos no perdidos y cuartiles. Pruebas: Wilcoxon de los rangos con signo, signo, McNemar. Si se ha instalado la opción Pruebas exactas (disponible sólo en los sistemas operativos Windows), la prueba de homogeneidad marginal también estará disponible.

Pruebas para dos muestras relacionadas: Consideraciones sobre los datos

Datos. Utilice variables numéricas que puedan ser ordenables.

Supuestos. Aunque no se suponen distribuciones en particular para las dos variables, se supone que la distribución de población de las diferencias emparejadas es simétrica.

Para obtener pruebas para dos muestras relacionadas

1. En los menús seleccione:

   Analizar > **Pruebas no paramétricas** > **Cuadros de diálogo antiguos** > **2 muestras relacionadas...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione uno o más pares de variables.

Dos muestras relacionadas: Tipos de pruebas

Las pruebas de esta sección comparan las distribuciones de dos variables relacionadas. La prueba apropiada depende del tipo de datos.

Si los datos son continuos, use la prueba de los signos o la prueba de Wilcoxon de los rangos con signo. La **prueba de los signos** calcula las diferencias entre las dos variables para todos los casos y clasifica las diferencias como positivas, negativas o empatadas. Si las dos variables tienen una distribución similar, el número de diferencias positivas y negativas no difiere de forma significativa. La **prueba de Wilcoxon de los rangos con signo** tiene en cuenta la información del signo de las diferencias y de la magnitud de las diferencias entre los pares. Dado que la prueba de Wilcoxon de los rangos con signo incorpora más información acerca de los datos, es más potente que la prueba de los signos.

Si los datos son binarios, use la **prueba de McNemar**. Esta prueba se utiliza normalmente en una situación de medidas repetidas, en la que la respuesta de cada sujeto se obtiene dos veces, una antes y otra después de que ocurra un evento especificado. La prueba de McNemar determina si el índice de respuesta inicial (antes del evento) es igual al índice de respuesta final (después del evento). Esta prueba es útil para detectar cambios en las respuestas causadas por la intervención experimental en los diseños del tipo antes-después.

Si los datos son categóricos, use la **prueba de homogeneidad marginal**. Se trata de una extensión de la prueba de McNemar a partir de la respuesta binaria a la respuesta multinomial. Contrasta los cambios de respuesta, utilizando la distribución chi-cuadrado, y es útil para detectar cambios de respuesta causados por intervención experimental en diseños antes-después. La prueba de homogeneidad marginal sólo está disponible si se ha instalado Pruebas exactas.

Pruebas para dos muestras relacionadas: Opciones

Estadísticos. Puede elegir uno o los dos estadísticos de resumen.

- **Descriptivos.** Muestra la media, la desviación estándar, el mínimo, el máximo y el número de casos no perdidos.
- **Cuartiles.** Muestra los valores correspondientes a los percentiles 25, 50 y 75.

Valores perdidos. Controla el tratamiento de los valores perdidos.

- **Excluir casos según prueba.** Cuando se especifican varias pruebas, cada una se evalúa separadamente respecto a los valores perdidos.
- **Excluir casos según lista.** Los casos con valores perdidos para cualquier variable se excluyen de todos los análisis.

Características adicionales del comando NPAR TESTS (Dos muestras relacionadas)

El lenguaje de sintaxis de comandos también permite contrastar una variable con cada variable de la lista.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Pruebas para varias muestras independientes

El procedimiento Pruebas para varias muestras independientes compara dos o más grupos de casos respecto a una variable.

Ejemplo. ¿Es diferente el tiempo medio en que se fundirán las bombillas de 100 vatios de tres marcas distintas? A partir del análisis de varianza de un factor de Kruskal-Wallis, puede comprobar que las tres marcas sí se diferencian en su vida media.

Estadísticas. Media, desviación estándar, mínimo, máximo, número de casos no perdidos y cuartiles. Pruebas: *H* de Kruskal-Wallis, de la mediana.

Pruebas para varias muestras independientes: Consideraciones sobre los datos

Datos. Utilice variables numéricas que puedan ser ordenables.

Supuestos. Utilice muestras independientes y aleatorias. La prueba *H* de Kruskal-Wallis requiere que las muestras comparadas tengan formas similares.

Para obtener pruebas para varias muestras independientes

1. En los menús seleccione:

Analizar > **Pruebas no paramétricas** > **Cuadros de diálogo antiguos** > **K muestras independientes...**

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione una o más variables numéricas.
3. Seleccione una variable de agrupación y pulse en **Definir rango** para especificar los valores enteros máximo y mínimo para la variable de agrupación.

Tipos de pruebas para varias muestras independientes

Se hallan disponibles tres pruebas para determinar si varias muestras independientes proceden de la misma población. La prueba *H* de Kruskal-Wallis, la prueba de la mediana y la prueba de Jonckheere-Terpstra contrastan si varias muestras independientes proceden de la misma población.

La prueba **H de Kruskal-Wallis**, una extensión de la prueba *U* de Mann-Whitney, es el análogo no paramétrico del análisis de varianza de un factor y detecta las diferencias en la localización de las distribuciones. La **prueba de la mediana**, que es una prueba más general pero no tan potente, detecta diferencias distribucionales en la localización y en la forma. La prueba *H* de Kruskal-Wallis y la prueba de la mediana suponen que no existe una ordenación *a priori* de las poblaciones *k* de las cuales se extraen las muestras.

Cuando *existe* una ordenación natural *a priori* (ascendente o descendente) de las poblaciones *k*, la prueba **Jonckheere-Terpstra** es más potente. Por ejemplo, las *k* poblaciones pueden representar *k* temperaturas ascendentes. Se contrasta la hipótesis de que diferentes temperaturas producen la misma distribución de respuesta, con la hipótesis alternativa de que cuando la temperatura aumenta, la magnitud de la respuesta aumenta. La hipótesis alternativa se encuentra aquí ordenada; por tanto, la prueba de Jonckheere-Terpstra es la prueba más apropiada. La prueba de Jonckheere-Terpstra estará disponible sólo si ha instalado el módulo adicional Pruebas exactas.

Varias muestras independientes: Definir rango

Para definir el rango, introduzca valores enteros para el **mínimo** y el **máximo** que se correspondan con las categorías mayor y menor de la variable de agrupación. Se excluyen los casos con valores fuera de los límites. Por ejemplo, si indica un valor mínimo de 1 y un valor máximo de 3, únicamente se utilizarán los valores enteros entre 1 y 3. Debe indicar ambos valores y el valor mínimo ha ser menor que el máximo.

Varias muestras independientes: Opciones

Estadísticos. Puede elegir uno o los dos estadísticos de resumen.

- **Descriptivos.** Muestra la media, la desviación estándar, el mínimo, el máximo y el número de casos no perdidos.
- **Cuartiles.** Muestra los valores correspondientes a los percentiles 25, 50 y 75.

Valores perdidos. Controla el tratamiento de los valores perdidos.

- **Excluir casos según prueba.** Cuando se especifican varias pruebas, cada una se evalúa separadamente respecto a los valores perdidos.
- **Excluir casos según lista.** Los casos con valores perdidos para cualquier variable se excluyen de todos los análisis.

Características adicionales del comando NPAR TESTS (K muestras independientes)

El lenguaje de sintaxis de comandos permite especificar un valor distinto al de la mediana observada para la prueba de la mediana (mediante el subcomando `MEDIAN`).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Pruebas para varias muestras relacionadas

El procedimiento Pruebas para varias muestras relacionadas compara las distribuciones de dos o más variables.

Ejemplo. ¿Asocia la gente diferentes niveles de prestigio a doctores, abogados, policías y profesores? Se pide a diez personas que ordenen estas cuatro profesiones por orden de prestigio. La prueba de Friedman indica que la gente asocia diferentes niveles de prestigio con estas cuatro profesiones.

Estadísticas. Media, desviación estándar, mínimo, máximo, número de casos no perdidos y cuartiles. Pruebas: Friedman, *W* de Kendall y *Q* de Cochran.

Pruebas para varias muestras relacionadas: Consideraciones sobre los datos

Datos. Utilice variables numéricas que puedan ser ordenables.

Supuestos. Las pruebas no paramétricas no requieren supuestos sobre la forma de la distribución subyacente. Utilice muestras aleatorias y dependientes.

Para obtener pruebas para varias muestras relacionadas

1. En los menús seleccione:

   Analizar > **Pruebas no paramétricas** > **Cuadros de diálogo antiguos** > **K muestras relacionadas...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione dos o más variables de contraste numéricas.

Tipos de prueba en el procedimiento Pruebas para varias muestras relacionadas

Hay tres pruebas disponibles para comparar las distribuciones de diversas variables relacionadas.

La **prueba de Friedman** es el equivalente no paramétrico de un diseño de medidas repetidas para una muestra o un análisis de varianza bidimensional con una observación por casilla. Friedman contrasta la hipótesis nula de que las *k* variables relacionadas procedan de la misma población. Para cada caso, las variables *k* se clasifican de 1 a *k*. La estadística de prueba se basa en estos rangos.

La **W de Kendall** es una normalización del estadístico de Friedman. La prueba *W* de Kendall se puede interpretar como el coeficiente de concordancia, que es una medida de acuerdo entre evaluadores. Cada caso es un juez o evaluador y cada variable es un elemento o persona que está siendo evaluada. Para cada variable, se calcula la suma de rangos. La *W* de Kendall varía entre 0 (no hay acuerdo) y 1 (acuerdo completo).

La prueba **Q de Cochran** es idéntica a la prueba de Friedman pero se puede aplicar cuando todas las respuestas son binarias. Esta prueba es una extensión de la prueba de McNemar para la situación de *k* muestras. La *Q* de Cochran contrasta la hipótesis de que diversas variables dicotómicas relacionadas tienen la misma media. Las variables se miden al mismo individuo o a individuos emparejados.

Pruebas para varias muestras relacionadas: Estadísticos

Puede elegir estadísticos.

- **Descriptivos.** Muestra la media, la desviación estándar, el mínimo, el máximo y el número de casos no perdidos.
- **Cuartiles.** Muestra los valores correspondientes a los percentiles 25, 50 y 75.

Características adicionales del comando NPAR TESTS (K muestras relacionadas)

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Análisis de respuestas múltiples

Análisis de respuestas múltiples

Se ofrecen dos procedimientos para analizar los conjuntos de categorías múltiples y de dicotomías múltiples. El procedimiento Frecuencias de respuestas múltiples muestra tablas de frecuencias. El procedimiento Tablas cruzadas de respuestas múltiples muestra tabulaciones cruzadas de dos y tres

dimensiones. Antes de utilizar cualquiera de los procedimientos, debe definir conjuntos de respuestas múltiples.

Ejemplo. Este ejemplo ilustra el uso de elementos de respuestas múltiples en un estudio de investigación de mercado. Los datos son ficticios y no deben interpretarse como reales. Una línea aérea podría hacer una encuesta a los pasajeros que realicen una determinada ruta para evaluar las líneas aéreas de la competencia. En este ejemplo, American Airlines desea conocer el uso que hacen sus pasajeros de otras líneas aéreas en la ruta Chicago-Nueva York y la importancia relativa del horario y el servicio a la hora de seleccionar una línea aérea. El encargado del vuelo proporciona a cada pasajero un breve cuestionario durante el embarque. La primera pregunta dice: rodee con un círculo todas las líneas aéreas con la que haya volado al menos una vez en los últimos seis meses en este mismo trayecto: American, United, TWA, USAir, Otras. Se trata de una pregunta de respuestas múltiples, ya que el pasajero puede marcar más de una respuesta. Sin embargo, la pregunta no se puede codificar directamente, ya que una variable sólo puede tener un valor para cada caso. Deberá utilizar distintas variables para correlacionar las respuestas con cada pregunta. Existen dos formas de hacerlo. Una consiste en definir una variable para cada una de las opciones (por ejemplo, American, United, TWA, USAir y Otras). Si el pasajero circula por United, a la variable *united* se le asigna un código de 1; de lo contrario, 0. Se trata de un **método de dicotomía múltiple** de correlación de variables. La otra forma de correlacionar respuestas es el **método de categorías múltiples**, en el que se estima el número máximo de posibles respuestas a la pregunta y se configura el mismo número de variables, con códigos para especificar la línea aérea utilizada. Examinando una muestra de cuestionarios, podría observarse que ningún usuario ha volado en más de tres líneas aéreas diferentes en esta ruta durante los últimos seis meses. Aún más, se observará que debido a la liberalización de las líneas aéreas, aparecen otras 10 en la categoría Otras. Con el método de respuestas múltiples, definiría tres variables, cada una codificada como 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta* y así sucesivamente. Si un pasajero determinado marca American y TWA, la primera variable tendrá el código 1, la segunda el 3 y la tercera un código de valor perdido. Otro pasajero podría haber marcado American e introducido Delta. Así, la primera variable tendrá el código 1, la segunda el 5 y la tercera un código de valor perdido. Por el contrario, si utiliza el método de dicotomías múltiples, terminará con 14 variables independientes. Aunque cualquiera de los métodos de correlación es factible para esta encuesta, el método que elija depende de la distribución de las respuestas.

Definir conjuntos de respuestas múltiples

El procedimiento Definir conjuntos de respuestas múltiples agrupa variables elementales en conjuntos de categorías múltiples y de dicotomías múltiples, para los que se pueden obtener tablas de frecuencias y tablas cruzadas. Se pueden definir hasta 20 conjuntos de respuestas múltiples. Cada conjunto debe tener un nombre exclusivo. Para eliminar un conjunto, resáltelo en la lista de conjuntos de respuestas múltiples y pulse **Borrar**. Para cambiar un conjunto, resáltelo en la lista, modifique cualquier característica de la definición del conjunto y pulse en **Cambiar**.

Las variables elementales se pueden codificar como dicotomías o categorías. Para utilizar variables dicotómicas, seleccione **Dicotomías** para crear un conjunto de dicotomías múltiples. Introduzca un valor entero en Valor contado. Cada variable que tenga al menos una aparición del valor contado se convierte en una categoría del conjunto de dicotomías múltiples. Seleccione **Categorías** para crear un conjunto de categorías múltiples con el mismo rango de valores que las variables que lo componen. Introduzca valores enteros para los valores máximo y mínimo del rango para las categorías del conjunto de categorías múltiples. El procedimiento suma cada valor entero distinto en el rango inclusivo para todas las variables que lo componen. Las categorías vacías no se tabulan.

A cada conjunto de respuestas múltiples se le debe asignar un nombre exclusivo de hasta siete caracteres. El procedimiento coloca delante del nombre asignado un signo dólar ($). No se pueden utilizar los siguientes nombres reservados: *casenum*, *sysmis*, *jdate*, *date*, *time*, *length* y *width*. El nombre del conjunto de respuestas múltiples sólo se encuentra disponible para su uso en los procedimientos de respuestas múltiples. No se puede hacer referencia a nombres de conjuntos de respuestas múltiples en otros procedimientos. Si lo desea, puede introducir una etiqueta de variable descriptiva para el conjunto de respuestas múltiples. La etiqueta puede tener hasta 40 caracteres.

Para definir conjuntos de respuestas múltiples

1. En los menús seleccione:

Analizar > **Respuesta múltiple** > **Definir conjuntos de variables...**

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione dos o más variables.
3. Si las variables están codificadas como dicotomías, indique qué valor desea contar. Si las variables están codificadas como categorías, defina el rango de las categorías.
4. Escriba un nombre exclusivo para cada conjunto de respuestas múltiples.
5. Pulse **Añadir** para añadir el conjunto de respuestas múltiples a la lista de conjuntos definidos.

Frecuencias de respuestas múltiples

El procedimiento Frecuencias de respuestas múltiples produce tablas de frecuencias para conjuntos de respuestas múltiples. En primer lugar es necesario definir uno o más conjuntos de respuestas múltiples (véase "Definir conjuntos de respuestas múltiples").

Para los conjuntos de dicotomías múltiples, los nombres de categorías que se muestran en los resultados proceden de etiquetas de variable definidas para variables elementales del grupo. Si las etiquetas de variable no están definidas, los nombres de las variables se utilizarán como etiquetas. Para los conjuntos de categorías múltiples, las etiquetas de categoría proceden de las etiquetas de valor de la primera variable del grupo. Si las categorías perdidas para la primera variable están presentes para otras variables del grupo, defina una etiqueta de valor para las categorías perdidas.

Valores perdidos. Los casos con valores perdidos se excluyen en base a tabla por tabla. Si lo desea, puede seleccionar una de las opciones siguientes o ambas:

- **Excluir los casos según lista dentro de las dicotomías.** Excluye los casos con valores perdidos en cualquier variable de la tabulación del conjunto de dicotomías múltiples. Esto sólo se aplica a conjuntos de respuestas múltiples definidos como conjuntos de dicotomías. De forma predeterminada, un caso se considera perdido para un conjunto de dicotomías múltiples si ninguna de sus variables que lo componen contiene el valor contado. Los casos con valores perdidos en algunas variables, pero no en todas, se incluyen en las tabulaciones del grupo si al menos una variable contiene el valor contado.
- **Excluir los casos según lista dentro de las categorías.** Excluye los casos con valores perdidos en cualquier variable de la tabulación del conjunto de categorías múltiples. Esto sólo se aplica a conjuntos de respuestas múltiples definidos como conjuntos de categorías. De forma predeterminada, un caso se considera perdido para un conjunto de categorías múltiples sólo si ninguno de sus componentes tiene valores válidos dentro del rango definido.

Ejemplo. Cada variable creada a partir de una pregunta de una encuesta es una variable elemental. Para analizar un elemento de respuestas múltiples, deberá combinar las variables en uno o dos tipos de conjuntos de respuestas múltiples: un conjunto de dicotomías múltiples o un conjunto de categorías múltiples. Por ejemplo, si una encuesta sobre líneas aéreas preguntara al encuestado cuál de las tres líneas (American, United, TWA) ha utilizado durante los seis últimos meses y usted utilizara variables dicotómicas y definiera un **conjunto de dicotomías múltiples**, cada una de las tres variables del conjunto se convertiría en una categoría de la variable de grupo. Las frecuencias y los porcentajes de las tres líneas aéreas se muestran en una tabla de frecuencias. Si observa que ningún encuestado ha mencionado más de dos líneas aéreas, podría crear dos variables, cada una con tres códigos, uno para cada línea aérea. Si define un **conjunto de categorías múltiples**, los valores se tabulan añadiendo los mismos códigos en las variables elementales juntas. El conjunto de valores resultantes es igual a los de cada una de las variables elementales. Por ejemplo, 30 respuestas para United son la suma de las cinco respuestas de United para la aerolínea 1 y las 25 respuestas de United para la aerolínea 2. Las frecuencias y los porcentajes de las tres líneas aéreas se muestran en una tabla de frecuencias.

Estadísticas. Tablas de frecuencias que muestran recuentos, porcentajes de respuestas, porcentajes de casos, número de casos válidos y número de casos perdidos.

Frecuencias de respuestas múltiples: Consideraciones sobre los datos

Datos. Utilice conjuntos de respuestas múltiples.

Supuestos. Las frecuencias y los porcentajes proporcionan una descripción útil de los datos de cualquier distribución.

Procedimientos relacionados. El procedimiento Definir conjuntos de respuestas múltiples permite definir este tipo de conjuntos.

Para obtener frecuencias de respuestas múltiples

1. En los menús seleccione:

   Analizar > **Respuesta múltiple** > **Frecuencias...**
2. Seleccione uno o más conjuntos de respuestas múltiples.

Tablas cruzadas de respuestas múltiples

El procedimiento Tablas cruzadas de respuestas múltiples presenta en forma de tabla cruzada conjuntos de respuestas múltiples, variables elementales o una combinación. También puede obtener porcentajes de casillas basados en casos o respuestas, modificar la gestión de los valores perdidos u obtener tabulaciones cruzadas emparejadas. Antes debe definir uno o varios conjuntos de respuestas múltiples (véase “Para definir conjuntos de respuestas múltiples”).

Para los conjuntos de dicotomías múltiples, los nombres de categorías que se muestran en los resultados proceden de etiquetas de variable definidas para variables elementales del grupo. Si las etiquetas de variable no están definidas, los nombres de las variables se utilizarán como etiquetas. Para los conjuntos de categorías múltiples, las etiquetas de categoría proceden de las etiquetas de valor de la primera variable del grupo. Si las categorías perdidas para la primera variable están presentes para otras variables del grupo, defina una etiqueta de valor para las categorías perdidas. El procedimiento muestra las etiquetas de categoría por columnas en tres líneas, con un máximo de ocho caracteres por línea. Para evitar la división de palabras, puede invertir los elementos de las filas y las columnas o volver a definir las etiquetas.

Ejemplo. Tanto los conjuntos de categorías múltiples como los conjuntos de dicotomías múltiples se pueden presentar en forma de tabla cruzada con otras variables de este procedimiento. Un estudio sobre pasajeros de líneas aéreas solicita a éstos la siguiente información: marque las líneas aéreas con las que ha volado al menos una vez en los seis últimos meses (American, United, TWA). ¿Qué considera más importante a la hora de seleccionar un vuelo, el horario o el servicio? Seleccione sólo uno. Después de introducir los datos como dicotomías o categorías múltiples y combinarlos en un conjunto, puede presentar en forma de tabla cruzada las selecciones de línea aérea con la pregunta relativa al servicio o al horario.

Estadísticas. Tabulación cruzada con recuentos de casilla, fila, columna y totales, así como porcentajes de casillas, filas, columnas y totales. Los porcentajes de casillas se basan en casos o respuestas.

Tablas cruzadas de respuestas múltiples: Consideraciones sobre los datos

Datos. Utilice conjuntos de respuestas múltiples o variables categóricas numéricas.

Supuestos. Las frecuencias y los porcentajes proporcionan una útil descripción de los datos de cualquier distribución.

Procedimientos relacionados. El procedimiento Definir conjuntos de respuestas múltiples permite definir este tipo de conjuntos.

Para obtener tablas cruzadas de respuestas múltiples

1. En los menús seleccione:

   Analizar > **Respuesta múltiple** > **Tablas cruzadas...**
2. Seleccione una o más variables numéricas o conjuntos de respuestas múltiples para cada dimensión de la tabla cruzada.
3. Defina el rango de cada variable elemental.

Si lo desea, puede obtener una tabla cruzada bidimensional para cada categoría de una variable de control o conjunto de respuestas múltiples. Seleccione uno o varios elementos para la lista Capas.

Tablas de respuestas múltiples: Definir rangos de las variables

Los rangos de valores deben definirse para cualquier variable elemental de la tabla cruzada. Introduzca los valores enteros de categoría máximos y mínimos que desee tabular. Las categorías que estén fuera del rango se excluyen del análisis. Se entiende que los valores que estén dentro del rango inclusivo son enteros (los no enteros quedan truncados).

Tablas cruzadas de respuestas múltiples: Opciones

Porcentajes de casillas. Los recuentos de casillas siempre se muestran. Puede elegir entre mostrar los porcentajes de fila, los de columna o los de tabla bidimensional (totales).

Porcentajes basados en. Los porcentajes de casillas pueden basarse en casos (o encuestados). Esta opción no estará disponible si selecciona la concordancia de variables en conjuntos de categorías múltiples. También se pueden basar en las respuestas. Para los conjuntos de dicotomías múltiples, el número de respuestas es igual al número de valores contados por los casos. Para los conjuntos de categorías múltiples, el número de respuestas es el número de valores del rango definido.

Valores perdidos. Puede elegir una o ambas de las siguientes opciones:

- **Excluir los casos según lista dentro de las dicotomías.** Excluye los casos con valores perdidos en cualquier variable de la tabulación del conjunto de dicotomías múltiples. Esto sólo se aplica a conjuntos de respuestas múltiples definidos como conjuntos de dicotomías. De forma predeterminada, un caso se considera perdido para un conjunto de dicotomías múltiples si ninguna de sus variables que lo componen contiene el valor contado. Los casos con valores perdidos para algunas variables, pero no todas, se incluyen en las tabulaciones del grupo si al menos una variable contiene el valor contado.
- **Excluir los casos según lista dentro de las categorías.** Excluye los casos con valores perdidos en cualquier variable de la tabulación del conjunto de categorías múltiples. Esto sólo se aplica a conjuntos de respuestas múltiples definidos como conjuntos de categorías. De forma predeterminada, un caso se considera perdido para un conjunto de categorías múltiples sólo si ninguno de sus componentes tiene valores válidos dentro del rango definido.

De forma predeterminada, cuando se presentan dos conjuntos de categorías múltiples en forma de tabla cruzada, el procedimiento tabula cada variable del primer grupo con cada variable del segundo y suma las frecuencias de cada casilla; de esta forma, algunas respuestas pueden aparecer más de una vez en una tabla. Puede seleccionar la opción siguiente:

Emparejar las variables entre los conjuntos de respuesta. Empareja la primera variable del primer grupo con la primera variable del segundo, y así sucesivamente. Si selecciona esta opción, el procedimiento basará los porcentajes de casillas en las respuestas en lugar de hacerlo en los encuestados. El emparejamiento no está disponible para conjuntos de dicotomías múltiples o variables elementales.

Características adicionales del comando MULT RESPONSE

La sintaxis de comandos también le permite:

- Obtener tablas de tabla cruzada con un máximo de cinco dimensiones (con el subcomando BY).
- Cambiar las opciones de formato de salida, incluyendo la supresión de etiquetas de valor (con el subcomando FORMAT).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Informes de los resultados

Informes de los resultados

Los listados de casos y los estadísticos descriptivos son herramientas básicas para estudiar y presentar los datos. Puede obtener listados de casos con el Editor de datos o el procedimiento Resumir, frecuencias y estadísticos descriptivos con el procedimiento Frecuencias, y estadísticos de subpoblación con el

procedimiento Medias. Cada uno utiliza un formato diseñado para que la información sea clara. Si desea visualizar la información en un formato diferente, los resúmenes de informes en filas y los resúmenes de informes en columnas le proporcionan el control que necesita sobre la presentación de datos.

Resúmenes de informes en filas

Resúmenes de informes en filas genera informes en los cuales se presentan distintos estadísticos de resumen en filas. También se encuentran disponibles listados de los casos, con o sin estadísticos de resumen.

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

Ejemplo. Una empresa con una cadena de tiendas registra los datos de sus empleados, incluyendo el salario, el cargo, la tienda y la sección en la que trabaja cada uno. Se podría generar un informe que proporcione los datos individuales de cada empleado (listado) desglosados por tienda y sección (variables de segmentación), con estadísticos de resumen (por ejemplo, el salario medio) por tienda, sección y sección dentro de cada tienda.

Columnas de datos. Muestra una lista de las variables del informe para las que desea obtener el listado de los casos o los estadísticos de resumen y controla el formato de presentación de las columnas de datos.

Salto de columna. Muestra una lista de las variables de segmentación opcionales que dividen el informe en grupos y controla los estadísticos de resumen y los formatos de presentación de Salto de columna. Si hay varias variables de segmentación, se creará un grupo distinto para cada una de las categorías de las variables de segmentación dentro de las categorías de la variable de segmentación anterior en la lista. Las variables de segmentación deben ser variables categóricas discretas que dividan los casos en un número limitado de categorías con sentido. Los valores individuales de cada variable de segmentación aparecen ordenados en una columna distinta situada a la izquierda de todas las columnas de datos.

Informe. Controla las características globales del informe, incluyendo los estadísticos de resumen globales, la presentación de los valores perdidos, la numeración de las páginas y los títulos.

Mostrar casos. Muestra los valores reales (o etiquetas de valor) de las variables de la columna de datos para cada caso. Esto genera un informe a modo de listado, que puede ser mucho más largo que un informe de resumen.

Vista previa. Muestra sólo la primera página del informe. Esta opción es útil para ver una vista previa del formato del informe sin tener que procesar el informe completo.

Los datos están ordenados. Para los informes con variables de segmentación, el archivo de datos se debe ordenar por los valores de estas variables antes de generar el informe. Si el archivo de datos ya está ordenado por estos valores, se puede ahorrar tiempo de procesamiento seleccionando esta opción. Esta opción es especialmente útil después de generar la vista previa de un informe.

Para obtener un informe de resumen: Estadísticos en filas

1. En los menús seleccione:

   Analizar > **Informes** > **Resúmenes de informes en filas...**
2. Seleccione una o más variables para las columnas de datos. En el informe se genera una columna para cada variable seleccionada.
3. Para los informes ordenados y mostrados por subgrupos, seleccione una o más variables para Salto de columnas.
4. Para los informes con estadísticos de resumen para los subgrupos definidos por las variables de segmentación, seleccione la variable de segmentación de la lista de variables de Salto de columna y pulse en **Resumen**, en la sección Salto de columna, para especificar las medidas del resumen.
5. Para los informes con estadísticos de resumen globales, pulse **Resumen** para especificar las medidas de resumen.

Formato de los saltos y de las columnas de datos del informe

Los cuadros de diálogo de formato controlan los títulos y el ancho de las columnas, la alineación del texto y la presentación de los valores de los datos o de las etiquetas de valor. El formato de las columnas de datos controla el formato de las columnas de datos situadas en la parte derecha de la página del informe. La opción Formato de salto controla el formato de los saltos de columna situadas en la parte izquierda.

Título de la columna. Para la variable seleccionada, controla el título de la columna. Los títulos largos se ajustan de forma automática dentro de la columna. Utilice la tecla Intro para insertar manualmente líneas de separación donde desee ajustar los títulos.

Posición de valor en la columna. Para la variable seleccionada, controla la alineación de los valores de los datos o de las etiquetas de valor dentro de la columna. La alineación de los valores o de las etiquetas no afecta a la alineación de los encabezados de las columnas. Puede sangrar el contenido de la columna por un número específico de caracteres o centrar el contenido.

Contenido de la columna. Para la variable seleccionada, controla la presentación de los valores de los datos o de las etiquetas de valor definidas. Los valores de los datos siempre se muestran para cualquier valor que no tenga etiquetas de valor definidas. No se encuentra disponible para las columnas de datos en los informes estadísticos en columnas.

Líneas de resumen finales y Líneas de resumen del informe

Los dos cuadros de diálogo Líneas de resumen controlan la presentación de los estadísticos de resumen para los grupos de ruptura y para el informe entero. Líneas de resumen controla los estadísticos de subgrupo para cada categoría definida por las variables de segmentación. Líneas de resumen finales controla los estadísticos globales que se muestran al final del informe.

Los estadísticos de resumen disponibles son: suma, media, valor mínimo, valor máximo, número de casos, porcentaje de casos por encima y por debajo de un valor especificado, porcentaje de casos dentro de un rango de valores especificado, desviación estándar, curtosis, varianza y asimetría.

Opciones de saltos del informe

Opciones de saltos controla el espaciado y la paginación de la información de la categoría de salto.

Control de página. Controla el espaciado y la paginación para las categorías de la variable de segmentación seleccionada. Puede especificar un número de líneas en blanco entre las categorías de segmentación o empezar cada categoría de segmentación en una página nueva.

Líneas en blanco antes de los estadísticos. Controla el número de líneas en blanco entre las etiquetas o los datos de la categoría de ruptura y los estadísticos de resumen. Esta opción es especialmente útil para los informes combinados que incluyan tanto el listado de los casos individuales como los estadísticos de resumen para las categorías de segmentación; en estos informes puede insertar un espacio entre el listado de los casos y los estadísticos de resumen.

Opciones del informe

Informe: Opciones controla el tratamiento y la presentación de los valores perdidos y la numeración de las páginas del informe.

Excluir casos con valores perdidos según lista. Elimina (del informe) cualquier caso con valores perdidos para cualquier variable del informe.

Los valores perdidos aparecen como. Permite especificar el símbolo que representa los valores perdidos en el archivo de datos. Este símbolo sólo puede tener un carácter y se utiliza para representar tanto los valores *perdidos del sistema* como los valores *perdidos del usuario*.

Numerar las páginas desde la. Permite especificar un número de página para la primera página del informe.

Diseño del informe

Informe: Diseño controla el ancho y alto de cada página del informe, la ubicación del informe dentro de la página y la inserción de etiquetas y líneas en blanco.

Diseño de página. Controla los márgenes de las páginas expresados en líneas (extremos superior e inferior) y caracteres (a la izquierda y a la derecha) y la alineación del informe entre los márgenes.

Títulos y pies de página. Controla el número de líneas que separan los títulos y los pies de página del cuerpo del informe.

Salto de columna. Controla la presentación de los saltos de columna. Si se especifican diversas variables de segmentación, pueden situarse en columnas diferentes o en la primera columna. Si se colocan todas en la primera columna, se generará un informe más estrecho.

Títulos de columna. Controla la presentación de los títulos de columna, incluyendo el subrayado de títulos, el espacio entre los títulos y el cuerpo del informe y la alineación vertical de los títulos de columna.

Filas de col. datos y etiquetas de salto. Controla la ubicación de la información de las columnas de datos (valores de datos o estadísticos de resumen) en relación con las etiquetas de salto al principio de cada categoría de ruptura. La primera fila de información puede empezar en la misma línea que la etiqueta de categoría de ruptura o en un número de líneas posterior especificado. Esta sección no se encuentra disponible para los informes de estadísticos en columnas.

Títulos del informe

Informe: Títulos controla el contenido y la ubicación de los títulos y los pies de página del informe. Puede especificar un máximo de diez líneas de títulos de página y otras tantas de pies de página, con componentes justificados a la izquierda, en el centro y a la derecha en cada línea.

Si inserta variables en los títulos o en los pies de página, la etiqueta de valor o el valor de la variable actual aparecerá en el título o en el pie de página. Para los títulos se mostrará la etiqueta de valor correspondiente al valor de la variable al principio de la página. Para los pies de página, esta etiqueta se mostrará al final de la página. Si no hay etiqueta de valor, se mostrará el valor real.

Variables especiales. Las variables especiales *DATE* y *PAGE* permiten insertar la fecha actual o el número de página en cualquier línea de un encabezado o pie del informe. Si el archivo de datos contiene variables llamadas *DATE* o *PAGE*, no podrá utilizar estas variables en los títulos ni en los pies del informe.

Resúmenes de informes en columnas

Resúmenes de informes en columnas genera informes de resumen en los que diversos estadísticos de resumen aparecen en columnas distintas.

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

Ejemplo. Una empresa con una cadena de tiendas registra la información de los empleados, incluyendo el salario, el cargo y la sección en la que trabaja cada uno. Se podría generar un informe que proporcione los estadísticos de salario resumidos (por ejemplo, media, mínimo y máximo) para cada sección.

Columnas de datos. Muestra una lista de las variables del informe para las que se desea obtener estadísticos de resumen y controla el formato de presentación y los estadísticos de resumen mostrados para cada variable.

Salto de columna. Muestra una lista de las variables de segmentación opcionales que dividen el informe en grupos y controla los formatos de presentación de los saltos de columna. Si hay varias variables de segmentación, se creará un grupo distinto para cada una de las categorías de las variables de segmentación dentro de las categorías de la variable de segmentación anterior en la lista. Las variables de segmentación deben ser variables categóricas discretas que dividan los casos en un número limitado de categorías con sentido.

Informe. Controla las características globales del informe, incluyendo la presentación de los valores perdidos, la numeración de las páginas y los títulos.

Vista previa. Muestra sólo la primera página del informe. Esta opción es útil para ver una vista previa del formato del informe sin tener que procesar el informe completo.

Los datos están ordenados. Para los informes con variables de segmentación, el archivo de datos se debe ordenar por los valores de estas variables antes de generar el informe. Si el archivo de datos ya está ordenado por estos valores, se puede ahorrar tiempo de procesamiento seleccionando esta opción. Esta opción es especialmente útil después de generar la vista previa de un informe.

Para obtener un informe de resumen: Estadísticos en columnas

1. En los menús seleccione:

   Analizar > **Informes** > **Resúmenes de informes en columnas...**
2. Seleccione una o más variables para las columnas de datos. En el informe se genera una columna para cada variable seleccionada.
3. Para cambiar la medida de resumen para una variable, seleccione la variable de la lista de columnas de datos y pulse en **Resumen**.
4. Para obtener más de una medida de resumen para una variable, seleccione la variable en la lista de origen y desplácela hasta la lista Columnas de datos varias veces, una para cada medida que desee obtener.
5. Para mostrar una columna con la suma, la media, la razón o cualquier otra función de las columnas existentes, pulse **Insertar total**. Al hacerlo se situará una variable llamada *total* en la lista Columnas de datos.
6. Para los informes ordenados y mostrados por subgrupos, seleccione una o más variables para Salto de columnas.

Función Columna de resumen total

Líneas de resumen controla el estadístico de resumen mostrado para la variable de las columnas de datos seleccionada.

Los estadísticos de resumen disponibles son: suma, media, valor mínimo, valor máximo, número de casos, porcentaje de casos por encima y por debajo de un valor especificado, porcentaje de casos dentro de un rango de valores especificado, desviación estándar, varianza, curtosis y asimetría.

Columna de resumen total

Columna de resumen controla los estadísticos de resumen del total que resumen dos o más columnas de datos.

Los estadísticos de resumen del total son la suma de columnas, la media de columnas, el mínimo, el máximo, la diferencia entre los valores de dos columnas, el cociente de los valores de una columna dividido por los valores de otra y el producto de los valores de las columnas multiplicados entre sí.

Suma de columnas. La columna *total* es la suma de las columnas de la lista Columna de resumen.

Media de columnas. La columna *total* es la media de las columnas de la lista Columna de resumen.

Mínimo de columnas. La columna *total* es el mínimo de las columnas de la lista Columna de resumen.

Máximo de columnas. La columna *total* es el máximo de las columnas de la lista Columna de resumen.

1ª columna - 2ª columna. La columna *total* es la resta de las columnas de la lista Columna de resumen. Esta lista debe contener, exactamente, dos columnas.

1ª columna / 2ª columna. La columna *total* es el cociente de las columnas de la lista Columna de resumen. Esta lista debe contener, exactamente, dos columnas.

% 1ª columna / 2ª columna. La columna *total* es el porcentaje de la primera columna dividido por la segunda columna de la lista Columna de resumen. Esta lista debe contener, exactamente, dos columnas.

Producto de columnas. La columna *total* es el producto de las columnas de la lista Columna de resumen.

Formato de columna de informe

Las opciones de formato de datos y salto de columna para Resúmenes de informes en columnas son iguales que las que se describen para Informes de estadísticos en filas.

Opciones de Salto de columna para los estadísticos en el informe

Opciones de saltos controla la presentación del subtotal, el espaciado y la paginación para las categorías de segmentación.

Subtotal. Controla los subtotales mostrados para cada categoría de segmentación.

Control de página. Controla el espaciado y la paginación para las categorías de la variable de segmentación seleccionada. Puede especificar un número de líneas en blanco entre las categorías de segmentación o empezar cada categoría de segmentación en una página nueva.

Líneas en blanco antes del subtotal. Controla el número de líneas en blanco entre los datos de las categorías de ruptura y los subtotales.

Opciones de columnas para los estadísticos en el informe

Opciones controla la presentación de los totales finales y de los valores perdidos y la paginación de los informes de estadísticos en columnas.

Total final. Muestra y etiqueta un total global para cada columna que aparece al final de la columna.

Valores perdidos. Permite excluir los valores perdidos del informe o seleccionar un único carácter para indicar estos valores.

Diseño de informe para Resúmenes de informes en columnas

Las opciones de diseño de informe para Resúmenes de informes en columnas son iguales que las que se describen para Resúmenes de informes en filas.

Características adicionales del comando REPORT

La sintaxis de comandos también le permite:

- Mostrar funciones de resumen diferentes en las columnas de una única línea de resumen.
- Insertar líneas de resumen en las columnas de datos para variables que no sean la variable de la columna de datos o para diversas combinaciones (funciones compuestas) de las funciones de resumen.
- Utilizar Mediana, Modo, Frecuencia y Porcentaje como funciones de resumen.
- Controlar de forma más precisa el formato de presentación de los estadísticos de resumen.
- Insertar líneas en blanco en diversos puntos de los informes.
- Insertar líneas en blanco después de cada e*n*-ésimo caso de los informes en formato de listado.

Debido a la complejidad de la sintaxis de REPORT, a la hora de generar un nuevo informe con sintaxis puede resultarle útil, para aproximar el informe generado a partir de los cuadros de diálogo, copiar y pegar la sintaxis correspondiente y depurar esa sintaxis para generar exactamente el informe que le interese.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Análisis de fiabilidad

El análisis de fiabilidad permite estudiar las propiedades de las escalas de medición y de los elementos que componen las escalas. El procedimiento Análisis de fiabilidad calcula un número de medidas de fiabilidad de escala que se utilizan normalmente y también proporciona información sobre las relaciones entre elementos individuales de la escala. Los coeficientes de correlación intraclase se pueden utilizar para calcular las estimaciones de fiabilidad entre evaluadores.

El análisis de fiabilidad también proporciona estadísticas de Kappa de varios evaluadores de Fleiss que evalúan el acuerdo entre evaluadores para determinar la fiabilidad entre los diversos evaluadores. Un acuerdo más alto proporciona más confianza en las puntuaciones que reflejan la verdadera circunstancia. Las opciones de Kappa de varios evaluadores de Fleiss están disponibles en el diálogo "Análisis de fiabilidad: Estadísticos" en la página 269.

Ejemplo

¿El cuestionario mide la satisfacción del cliente de manera útil? El análisis de fiabilidad le permitirá determinar el grado en que los elementos del cuestionario se relacionan entre sí, obtener un índice global de la replicabilidad o de la consistencia interna de la escala en su conjunto e identificar elementos problemáticos que deberían ser excluidos de la escala.

Estadísticos

Descriptivos para cada variable y para la escala, estadísticas de resumen entre elementos, correlaciones y covarianzas entre elementos, estimaciones de fiabilidad, tabla ANOVA, coeficiente de correlación intraclase, T^2 de Hotelling, prueba de aditividad de Tukey y Kappa de varios evaluadores de Fleiss.

Modelos

Están disponibles los siguientes modelos de fiabilidad:

Alfa (Cronbach)

Este modelo es una medida de coherencia interna basada en el promedio de correlación entre elementos.

Omega (de McDonald)

Este modelo presupone que el modelo es unidimensional e incluye un único factor sin dependencia de elementos locales en forma de covarianzas de error. El modelo implica que la covarianza de los dos elementos distintos es el producto de sus cargas.

Dos mitades

Este modelo divide la escala en dos partes y examina la correlación entre dichas partes.

Guttman

Este modelo calcula los límites inferiores de Guttman para la fiabilidad verdadera.

Paralelo

Este modelo asume que todos los elementos tienen varianzas iguales y varianzas error iguales a través de las réplicas.

Paralelo estricto

Este modelo asume los supuestos del modelo paralelo y también asume que las medias son iguales a través de los elementos.

Consideraciones sobre datos de análisis de fiabilidad

Datos

Los datos pueden ser dicotómicos, ordinales o de intervalo, pero deben estar codificados numéricamente.

Supuestos

Las observaciones deben ser independientes y los errores no deben estar correlacionados entre los elementos. Cada par de elementos debe tener una distribución normal bivariada. Las escalas deben ser aditivas, de manera que cada elemento esté linealmente relacionado con la puntuación total. Los siguientes supuestos se aplican para las estadísticas de Kappa de varios evaluadores de Fleiss:

- Se deben seleccionar al menos dos variables de elemento para ejecutar cualquier estadística de fiabilidad.
- Cuando se seleccionan al menos dos variables de puntuación, se pega la sintaxis de Kappa de varios evaluadores de Fleiss.
- No hay ninguna conexión entre los evaluadores.
- El número de evaluadores es una constante.
- Cada sujeto está clasificado por el mismo grupo que contiene únicamente un solo evaluador.
- No se pueden asignar ponderaciones a los diversos desacuerdos.

Procedimientos relacionados

Si desea explorar la dimensionalidad de los elementos de la escala (para comprobar si es necesario más de un constructor para explicar el patrón de puntuaciones en los elementos), utilice el Análisis factorial o el Escalamiento multidimensional. Para identificar grupos homogéneos de variables, use el análisis de clústeres jerárquico para agrupar las variables en clústeres.

Para obtener un análisis de fiabilidad

1. En los menús seleccione:

   Analizar > **Escala** > **Análisis de fiabilidad...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione dos o más variables como componentes potenciales de una escala aditiva.
3. Elija un modelo en la lista desplegable **Modelo**.
4. Si lo desea, pulse **Estadísticos** para seleccionar varias estadísticas que describen los elementos de la escala o el acuerdo entre evaluadores.

Análisis de fiabilidad: Estadísticos

Puede seleccionar diversas estadísticas que describan la escala, los elementos y el acuerdo entre evaluadores para determinar la fiabilidad entre los diversos evaluadores. Los estadísticos de los que se informa de forma predeterminada incluyen el número de casos, el número de elementos y las estimaciones de la fiabilidad, según se explica a continuación:

Modelos alfa

Coeficiente Alfa; para datos dicotómicos, éste es equivalente al coeficiente 20 de Kuder-Richardson (KR20).

Modelos Omega

Estimación de omega de McDonald para evaluar la fiabilidad.

Modelos de dos mitades

Correlación entre formas, fiabilidad de dos mitades de Guttman, fiabilidad de Spearman-Brown (longitud igual y desigual) y coeficiente alfa para cada mitad.

Modelos Guttman

Coeficientes de fiabilidad lambda 1 a lambda 6.

Modelos paralelos y estrictamente paralelos

Prueba de bondad de ajuste del modelo; estimaciones de la varianza error, varianza común y varianza verdadera; correlación común entre elementos estimada; fiabilidad estimada y estimación de la fiabilidad no sesgada.

Descriptivos para

Genera estadísticos descriptivos para las escalas o los elementos a través de los casos.

Elemento

Genera estadísticos descriptivos para los elementos a través de los casos.

Escala
Genera estadísticos descriptivos para las escalas.

Escala si se elimina el elemento
Muestra estadísticos de resumen para comparar cada elemento con la escala compuesta por otros elementos. Los estadísticos incluyen la media de escala y la varianza si el elemento fuera a eliminarse de la escala, la correlación entre el elemento y la escala compuesta por otros elementos, y alfa de Cronbach si el elemento fuera a eliminarse de la escala.

Resúmenes
Proporciona estadísticos descriptivos sobre las distribuciones de los elementos a través de todos los elementos de la escala.

Medias
Estadísticos de resumen para las medias de los elementos. Se muestran el máximo, el mínimo y el promedio de las medias de los elementos, el rango y la varianza de las medias de los elementos, y la razón de la mayor media sobre la menor media de los elementos.

Varianzas
Estadísticos de resumen para las varianzas de los elementos. Se muestran el máximo, el mínimo y el promedio de las varianzas de los elementos, el rango y la varianza de las varianzas de los elementos y la razón de la mayor varianza sobre la menor varianza de los elementos.

Correlaciones
Estadísticos de resumen para correlaciones entre elementos. Se muestran el máximo, el mínimo y el promedio de las correlaciones entre elementos, el rango y la varianza de las correlaciones entre elementos, y la razón de la mayor correlación sobre la menor correlación entre elementos.

Covarianzas
Estadísticos de resumen para las covarianzas entre elementos. Se muestran el máximo, el mínimo y el promedio de las covarianzas entre elementos, el rango y la varianza de las covarianzas entre elementos, y la razón de la mayor sobre la menor covarianza entre elementos.

Inter-elementos
Genera las matrices de correlaciones o covarianzas entre los elementos.

Tabla de ANOVA
Produce pruebas de medias iguales.

Prueba F
Muestra la tabla de un análisis de varianza de medidas repetidas.

Chi-cuadrado de Friedman
Muestra el chi-cuadrado de Friedman y el coeficiente de concordancia de Kendall. Esta opción es adecuada para datos que se encuentren en el formato de rangos. La prueba de chi-cuadrado sustituye a la prueba F habitual en la tabla de ANOVA.

Chi-cuadrado de Cochran
Muestra Q de Cochran. Esta opción es adecuada para los datos que son dicotómicos. El estadístico Q sustituye a la F habitual en la tabla de ANOVA.

Acuerdo entre evaluadores: Kappa de Fleiss
Evalúa el acuerdo entre evaluadores para determinar la fiabilidad entre los diversos evaluadores. Un acuerdo más alto proporciona más confianza en las puntuaciones que reflejan la verdadera circunstancia. Las estadísticas kappa no ponderadas generalizadas miden el acuerdo entre cualquier número de constante de evaluadores, asumiendo que:

- Se deben especificar al menos dos variables de elemento para ejecutar cualquier estadística de fiabilidad.
- Se deben especificar al menos dos variables de puntuación.
- Las variables seleccionadas como elementos también se pueden seleccionar como puntuaciones.
- No hay ninguna conexión entre los evaluadores.
- El número de evaluadores es una constante.
- Cada sujeto está clasificado por el mismo grupo que contiene únicamente un solo evaluador.

- No se pueden asignar ponderaciones a los diversos desacuerdos.

Mostrar acuerdo sobre categorías individuales
Especifica si se va a generar o no el acuerdo sobre categorías individuales. De forma predeterminada, el resultado suprime la estimación en cualquier categoría individual. Cuando está habilitado, se muestran varias tablas en los resultados.

Ignorar mayúsculas/minúsculas
Controla si las variables de cadena son sensibles a mayúsculas y minúsculas. De forma predeterminada, los valores de puntuación de cadenas son sensibles a mayúsculas y minúsculas.

Las etiquetas de categoría de cadena se muestran en mayúsculas
Controla si las etiquetas de categoría de las tablas de resultados se muestran en mayúsculas o en minúsculas. El valor está habilitado de forma predeterminada, que muestra las etiquetas de categoría de cadena en mayúsculas.

Nivel de significación asintótica (%)
Especifica el nivel de significación para los intervalos de confianza asintótica. 95 es el valor predeterminado.

Perdido

Excluir tanto los valores perdidos del usuario como los valores perdidos del sistema
Controla la exclusión de los valores perdidos del usuario y perdidos del sistema. De forma predeterminada, se excluyen los valores perdidos del usuario y perdidos del sistema.

Los valores perdidos del usuario se tratan como válidos
Cuando está habilitado, trata a los valores perdidos del usuario y del sistema como datos válidos. De forma predeterminada, el valor está inhabilitado.

T-cuadrado de Hotelling
Genera una prueba multivariante sobre la hipótesis nula de que todos los elementos de la escala tienen la misma media.

Prueba de aditividad de Tukey
Genera una prueba sobre el supuesto de que no hay ninguna interacción multiplicativa entre los elementos.

Coeficiente de correlación intraclase
Genera medidas sobre la consistencia o sobre el acuerdo de los valores entre los propios casos.

Modelo
Seleccione el modelo para calcular el coeficiente de correlación intraclase. Los modelos disponibles son: Mixto de dos factores, aleatorio de dos factores y aleatorio de un factor. Seleccione **Mixto de dos factores**, si los efectos de personas son aleatorios y los efectos de elementos son fijos, **Aleatorio de dos factores** si los efectos de personas y los efectos de elementos son aleatorios; o **Aleatorio de un factor** si los efectos de personas son aleatorios.

Tipo
Seleccione el tipo de índice. Los tipos disponibles son: Consistencia y Acuerdo absoluto.

Intervalo de confianza (%)
Especifica el nivel para el intervalo de confianza. El valor predeterminado es 95%.

Valor de prueba
Especifica el valor hipotetizado para el coeficiente, para el contraste de hipótesis. Este valor es el valor con el que se compara el valor observado. El valor predeterminado es 0.

Características adicionales del comando RELIABILITY

La sintaxis de comandos también le permite:

- Leer y analizar una matriz de correlaciones.
- Escribir una matriz de correlaciones para su análisis posterior.
- Especificar divisiones que no sean mitades iguales para el método de dos mitades.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Kappa ponderada

La kappa ponderada de Cohen se utiliza ampliamente en la clasificación cruzada como medida del acuerdo entre evaluadores observados. Es un índice de acuerdo adecuado cuando las puntuaciones son escalas nominales sin estructura de orden. El desarrollo de kappa ponderada de Cohen se ha motivado por el hecho de que algunas asignaciones en una tabla de contingencia pueden tener mayor gravedad que otras. El estadístico se basa en las ponderaciones de las casillas predefinidas que reflejan un acuerdo o un desacuerdo.

El procedimiento de Kappa ponderada proporciona opciones para estimar la kappa ponderada de Cohen, una generalización importante del estadístico de kappa que mide el acuerdo de dos sujetos ordinales con categorías idénticas.

Nota: El procedimiento Kappa ponderado sustituye a la funcionalidad proporcionada anteriormente por la extensión `STATS WEIGHTED KAPPA.spe`.

Ejemplo

Hay situaciones en las que las diferencias entre los evaluadores no se deben tratar tan importantes. Un ejemplo sería en el sector sanitario en el que varias personas recogen datos de investigación o clínicos. En estos casos, la fiabilidad de los datos puede entrar en duda, dada la variabilidad entre los que recogen datos.

Estadísticos

Kappa ponderada de Cohen, escala lineal, escala cuadrática, intervalo de confianza asintótico.

Consideraciones sobre los datos de Kappa ponderada

Datos

Es necesaria una tabla bidireccionada que se basa en un conjunto de datos activo para estimar el estadístico de kappa ponderada de Cohen.

Las variables de puntuación deben ser del mismo tipo (toda una cadena o todo numéricos).

La estimación de kappa ponderada de Cohen sólo tiene sentido cuando las categorías de las dos variables de puntuación, representadas por la fila y la columna en la tabla, se ordenan de forma apropiada (para un par de variables numéricas, se aplica el orden numérico; para un par de variables de cadena, se aplica el orden alfabético).

Supuestos

En el caso de que se seleccionen pares de variables mixtas, kappa ponderada de Cohen no se estima.

Se supone que las variables de puntuación comparten el mismo conjunto de categorías.

Para obtener un análisis Kappa ponderada

1. En los menús seleccione:

   Analizar > **Escala** > **Kappa ponderada...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione dos o más variables numéricas o de cadena para especificar como **Evaluadores por parejas**.

   Nota: Debe seleccionar todas las variables de cadena o todas las variables numéricas.
3. Si lo desea, habilite el valor **Especificar evaluadores para filas y columnas** para controlar la visualización de los evaluadores por parejas o de los evaluadores de filas/columnas.
   - Cuando estén habilitados, los evaluadores por parejas se suprimen y se muestran los evaluadores de fila/columna. Las actualizaciones de la interfaz de usuario para proporcionar campos **Evaluadores de fila** y **Evaluadores de columna** (reemplazando de forma eficaz el campo **Evaluadores por parejas**).

- Cuando estén inhabilitados, los evaluadores de fila/columna se suprimen y se visualizan los evaluadores por parejas (el valor predeterminado)

Cuando **Especificar evaluadores para filas y columnas** esté habilitado, especifique como mínimo una variable para **Evaluadores de fila** y **Evaluadores de columna**.

Nota: Si los **Evaluadores de fila** y los **Evaluadores de columna** solo contienen una variable, las variables seleccionadas no pueden ser las mismas para ambos.

4. Si lo desea, pulse **Criterios** para especificar la escala de ponderación y los ajustes de valores que faltan o **Imprimir** para especificar el formato de visualización y los ajustes de tabla cruzada.

Kappa ponderada: Criterios

El diálogo Criterios proporciona opciones para especificar la estimación de los estadísticos de kappa ponderada de Cohen.

Escala de ponderación
Proporciona opciones para especificar una ponderación lineal o cuadrática para el acuerdo. El uso de la ponderación lineal (también conocida como Cicchetti-Allison) es el valor predeterminado.

Valores perdidos
Proporciona opciones para eliminar casos con valores perdidos en una base por parejas y para tratar los valores perdidos del usuario como válidos.

Las variables de puntuación de cadena distinguen entre mayúsculas y minúsculas
Cuando se selecciona, las variables de cadena se consideran como si distinguieran mayúsculas y minúsculas.

Intervalo de confianza asintótica (%)
Este valor opcional especifica el nivel de confianza para la estimación de los intervalos de confianza asintótica. Debe ser un único valor doble entre 0 y 100 (95 es el valor predeterminado).

Kappa ponderada: Imprimir

El diálogo Imprimir proporciona opciones para controlar las tablas cruzadas.

Mostrar y Dar formato
Proporciona opciones para controlar la visualización y el formato de la tabla cruzada.

Las categorías de puntuación se muestran en orden ascendente
Cuando se selecciona, las categorías de puntuación en las tablas cruzadas se muestran en orden ascendente. Cuando no se selecciona esta opción, las categorías de puntuación en las tablas cruzadas se muestran en orden descendente. Este valor está habilitado de forma predeterminada.

Las etiquetas de categoría de cadena se muestran en mayúsculas
Cuando se selecciona, las tablas cruzadas se muestran como mayúsculas. Cuando no se selecciona el valor, las tablas cruzadas, las tablas cruzadas se muestran en minúsculas. Este valor está habilitado de forma predeterminada.

Tabla cruzada
Proporciona opciones para especificar las variables de puntuación que se utilizan en la tabla cruzada. De forma predeterminada, los valores de tabla cruzada no están habilitados, lo que suprime la tabla cruzada de cualquier variable de puntuación.

Mostrar las referencias cruzadas en tablas de las variables de puntuación
Cuando se selecciona, este valor habilita la tabla cruzada de todas las variables de puntuación especificadas por el usuario o de todas las especificadas por el usuario.

Incluir todas las variables de puntuación
Cuando se selecciona, las tablas cruzadas se imprimen para todos los pares de variables de puntuación definidas.

Incluir variables de puntuación especificadas por usuario
Cuando se selecciona, utilice los campos **Variables disponibles**, **Evaluadores de fila** y **Evaluadores de columna** para seleccionar qué variables de puntuación se incluyen en las tablas cruzadas.

Escalamiento multidimensional

El escalamiento multidimensional trata de encontrar la estructura de un conjunto de medidas de distancia entre objetos o casos. Esta tarea se logra asignando las observaciones a posiciones específicas en un espacio conceptual (normalmente de dos o tres dimensiones) de modo que las distancias entre los puntos en el espacio concuerden al máximo con las disimilaridades dadas. En muchos casos, las dimensiones de este espacio conceptual son interpretables y se pueden utilizar para comprender mejor los datos.

Si las variables se han medido objetivamente, puede utilizar el escalamiento multidimensional como técnica de reducción de datos (el procedimiento Escalamiento multidimensional permitirá calcular las distancias a partir de los datos multivariantes, si es necesario). El escalamiento multidimensional puede también aplicarse a puntuaciones subjetivas de disimilaridad entre objetos o conceptos. Además, el procedimiento Escalamiento multidimensional puede tratar datos de disimilaridad procedentes de múltiples fuentes, como podrían ser múltiples evaluadores o múltiples encuestados.

Ejemplo. ¿Cómo percibe el público las diferencias entre distintos coches? Si posee datos de las puntuaciones de similaridad emitidas por los encuestados sobre las diferentes marcas y modelos de coches, puede utilizar el escalamiento multidimensional para identificar las dimensiones que describan las preferencias de los consumidores. Puede encontrar, por ejemplo, que el precio y el tamaño de un vehículo definen un espacio de dos dimensiones, capaz de explicar las similaridades de las que informan los encuestados.

Estadísticas. Para cada modelo: Matriz de datos, Matriz de datos escalada óptimamente, S-stress (de Young), estrés (de Kruskal), R^2, Coordenadas de los estímulos, estrés promedio y R^2 para cada estímulo (modelos RMDS). Para modelos de diferencias individuales (INDSCAL): ponderaciones del sujeto e índice de peculiaridad para cada sujeto. Para cada matriz en los modelos de escalamiento multidimensional replicado: estrés y R^2 para cada estímulo. Gráficos: coordenadas de los estímulos (de dos o tres dimensiones), diagrama de dispersión de las disparidades frente a las distancias.

Escalamiento multidimensional: Consideraciones sobre los datos

Datos. Si los datos son de disimilaridad, todas las disimilaridades deben ser cuantitativas y deben estar medidas en la misma métrica. Si los datos son datos multivariantes, las variables pueden ser datos cuantitativos, binarios o de recuento. El escalamiento de las variables es un tema importante, ya que las diferencias en el escalamiento pueden afectar a la solución. Si las variables tienen grandes diferencias en el escalamiento (por ejemplo, una variable se mide en dólares y otra en años), debe considerar la posibilidad de tipificarlas (este proceso puede llevarse a cabo automáticamente con el propio procedimiento Escalamiento multidimensional).

Supuestos. El procedimiento Escalamiento multidimensional está relativamente libre de supuestos distribucionales. Compruebe que selecciona el nivel de medición adecuado (ordinal, de intervalo, o de razón) en el cuadro de diálogo Escalamiento multidimensional: Opciones para asegurar que los resultados se calculan correctamente.

Procedimientos relacionados. Si su objetivo es la reducción de los datos, un método alternativo a tener en cuenta es el análisis factorial, sobre todo si las variables son cuantitativas. Si desea identificar grupos de casos similares, considere complementar el análisis de escalamiento multidimensional con un análisis de clústeres jerárquico o de *k*-medias.

Para obtener un análisis de escalamiento multidimensional

1. En los menús seleccione:

   Analizar > **Escala** > **Escalamiento multidimensional...**
2. Seleccione al menos cuatro variables para su análisis.
3. En el grupo Distancias, seleccione **Los datos son distancias** o **Crear distancias a partir de datos**.

4. Si selecciona **Crear distancias a partir de datos**, también podrá seleccionar una variable de agrupación para matrices individuales. La variable de agrupación puede ser numérica o de cadena.

Si lo desea, tiene la posibilidad de:

- Especifique la forma de la matriz de distancias cuando los datos sean distancias.
- Especifique la medida de distancia que hay que utilizar al crear distancias a partir de datos.

Uso del mapeo de proximidad para el escalado multidimensional

Además de ALSCAL, IBM SPSS Estadística ofrece **Proximity Mapping (PROXMAP)** como un método relacionado con el escalado multidimensional. ALSCAL y PROXMAP difieren en la forma en que se calcula el ajuste del modelo. ALSCAL ajusta las distancias al cuadrado a las disimilitudes al cuadrado, mientras que PROXMAP ajusta las distancias directamente a las disimilitudes. Por lo tanto, ambos procedimientos pueden arrojar valores numéricos distintos, incluso cuando las configuraciones espaciales resultantes parezcan similares. Estas diferencias pueden resultar más evidentes cuando se trabaja con datos que no tienen una solución casi perfecta.

Mientras que ALSCAL ofrece un escalado multidimensional básico con resultados textuales limitados y una información de diagnóstico mínima, PROXMAP ofrece soluciones modernas y completas con visualizaciones relativamente precisas (incluidos diagramas de Shepard, gráficos de transformación y análisis de residuos), resultados estadísticos mejorados (múltiples medidas de tensión, valores de R cuadrado e índices de ajuste detallados) y funciones avanzadas como el ajuste de propiedades.

Consulte los siguientes enlaces para empezar a utilizar Proximity Mapping:

- “Comparación entre ALSCAL y PROXMAP” en la página 276
- Correlación de proximidad
-
-
-

Escalamiento multidimensional: Forma de los datos

Si el conjunto de datos activo representa distancias entre uno o dos conjuntos de objetos, especifique la forma de la matriz de datos para obtener los resultados correctos.

Nota: no puede seleccionar **Cuadrada simétrica** si el cuadro de diálogo Modelo especifica la condicionalidad de filas.

Escalamiento multidimensional: Crear la medida a partir de los datos

El escalamiento multidimensional utiliza datos de disimilaridad para crear una solución de escalamiento. Si los datos son datos multivariantes (los valores de las variables que se han medido), debe crear los datos de disimilaridad para poder calcular una solución de escalamiento multidimensional. Puede especificar los detalles para la creación de las medidas de disimilaridad a partir de los datos.

Medida. Le permite especificar la medida de disimilaridad para el análisis. Seleccione una opción del grupo Medida que se corresponda con el tipo de datos y, a continuación, elija una de las medidas de la lista desplegable correspondiente a ese tipo de medida. Las opciones disponibles son:

- **Intervalo**. Distancia euclídea, Distancia euclídea al cuadrado, Chebychev, Bloque, Minkowski o Personalizada.
- **Contar apariciones**. Medida de chi-cuadrado o Medida de phi-cuadrado.
- **Binario**. Distancia euclídea, Distancia euclídea al cuadrado, Diferencia de tamaño, Diferencia de configuración, Varianza o Lance y Williams.

Crear matriz de proximidades. Le permite elegir la unidad de análisis. Las opciones son Entre variables o Entre casos.

Transformar valores. En determinados casos, como cuando las variables se miden en escalas muy distintas, puede que desee tipificar los valores antes de calcular las proximidades (no es aplicable a datos binarios). Seleccione un método de estandarización en la lista desplegable Estandarizar. Si no se requiere ninguna estandarización, seleccione **Ninguno**.

Escalamiento multidimensional: Modelo

La estimación correcta de un modelo de escalamiento multidimensional depende de aspectos que atañen a los datos y al modelo en sí.

Nivel de medición. Permite especificar el nivel de los datos. Las opciones son Ordinal, Intervalo y Razón. Si las variables son ordinales, al seleccionar **Desempatar observaciones empatadas** se solicitará que sean consideradas como variables continuas, de forma que los empates (valores iguales para casos diferentes) se resuelvan óptimamente.

Condicionalidad. Permite especificar qué comparaciones tienen sentido. Las opciones son Matriz, Fila o Incondicional.

Dimensiones. Permite especificar la dimensionalidad de la solución o soluciones del escalamiento. Se calcula una solución para cada número del rango especificado. Especifique números enteros entre 1 y 6; se permite un mínimo de 1 sólo si selecciona **Distancia euclídea** como modelo de escalamiento. Para una solución única, especifique el mismo número para el mínimo y el máximo.

Modelo de escalamiento. Permite especificar los supuestos bajo los que se realiza el escalamiento. Las opciones disponibles son Distancia euclídea o Distancia euclídea de diferencias individuales (también conocida como INDSCAL). Para el modelo de Distancia euclídea de diferencias individuales, puede seleccionar **Permitir ponderaciones negativas de los sujetos**, si es adecuado para los datos.

Escalamiento multidimensional: Opciones

Puede especificar opciones para el análisis de escalamiento multidimensional.

Representación. Permite seleccionar varios tipos de resultados. Las opciones disponibles son Gráficos de grupo, Gráficos para los sujetos individuales, Matriz de datos y Resumen del modelo y de las opciones.

Criterios. Permite determinar cuándo debe detenerse la iteración. Para cambiar los valores predeterminados, introduzca valores para la **Convergencia de s-stress**, el **Valor mínimo de s-stress** y el **Nº máximo de iteraciones**.

Tratar distancias menores que n como perdidas. Las distancias menores que este valor se excluyen del análisis.

Características adicionales del comando ALSCAL

La sintaxis de comandos también le permite:

- Utilizar tres tipos de modelos adicionales, conocidos como ASCAL, AINDS y GEMSCAL en la literatura sobre escalamiento multidimensional.
- Realizar transformaciones polinómicas en los datos de intervalo y proporción.
- Analizar similaridades (en lugar de distancias) con datos ordinales.
- Analizar datos nominales.
- Guardar varias matrices de coordenadas y ponderación en archivos y volverlas a leer para análisis.
- Restringir el desplegamiento multidimensional.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Comparación entre ALSCAL y PROXMAP

Tanto ALSCAL como PROXMAP realizan un escalamiento multidimensional (MDS) para representar los datos de proximidad en un número determinado de dimensiones. En PROXMAP, estos resultados

se presentan utilizando una terminología actualizada y gráficos de diagnóstico que permiten una comparación directa con los resultados de ALSCAL.

ALSCAL y PROXMAP calculan la disposición espacial de los objetos basándose en datos de proximidad. PROXMAP ofrece un marco basado en ventanas de diálogo que admite los mismos resultados conceptuales que ALSCAL, utilizando terminología actualizada y gráficos de diagnóstico consolidados.

Mapas conceptuales

La siguiente tabla muestra la correspondencia entre los términos y resultados clave de ALSCAL y la terminología de PROXMAP. El significado analítico de los resultados es el mismo en todos los procedimientos.

Tabla 5. Correspondencia terminológica entre ALSCAL y PROXMAP

Término o resultado de ALSCAL	**Equivalente a PROXMAP**
Configuración de estímulo derivado	Espacio común de objetos
Diagrama de dispersión o ajuste lineal Disparidades Distancias	**Fit Source <Etiqueta de origen>** Proximidades transformadas Distancias
Diagrama de dispersión o ajuste no lineal Observaciones Distancias	**Gráfico de Shepard**[1]**Fuente <Etiqueta de fuente>** Proximidades originales Distancias + Proximidades transformadas
Diagrama de dispersión de transformación Observaciones Disparidades	**Fuente de transformación <Etiqueta de la fuente>** Proximidades originales Proximidades transformadas

Pasar de ALSCAL a PROXMAP

El siguiente tutorial muestra cómo se puede reproducir un análisis de escalamiento multidimensional (MDS) de ALSCAL utilizando PROXMAP, y explica cómo interpretar las similitudes y diferencias en los resultados.

Visión general

En este tutorial se muestra cómo se configura PROXMAP para replicar un análisis de ALSCAL y cómo se pueden comparar las configuraciones resultantes.

El ejemplo se centra en un único conjunto de datos de proximidad y utiliza los ajustes predeterminados o equivalentes, de modo que las diferencias entre ALSCAL y PROXMAP se limitan al escalado y la orientación, en lugar de a cambios sustanciales en la solución.

Datos utilizados en el análisis

El análisis se basa en Dist_Cities_LowerTr, un conjunto de datos de proximidad que recoge las distancias de vuelo entre doce ciudades de Estados Unidos. Los datos se almacenan como una matriz de proximidad simétrica en forma triangular inferior.

[1] El término «diagrama de Shepard» hace referencia a Roger Shepard, uno de los creadores del escalamiento multidimensional no métrico. Combina el gráfico «Ajuste no lineal» de ALSCAL y el gráfico «Transformación».

En la vista de datos, el conjunto de datos representa una matriz de proximidad simétrica de tipo « 12x12 » almacenada en formato triangular inferior y una variable de identificador de objeto nominal utilizada para etiquetar la configuración.

En la vista «Variable», el conjunto de datos está compuesto por doce variables de nivel de escala que contienen la matriz de proximidad y una variable nominal que define las etiquetas de los objetos.

La variable *CITY* se especifica como la variable de etiqueta del objeto.

Además de la matriz de proximidad, el conjunto de datos incluye una variable de identificador nominal de objetos (*CITY*) que proporciona etiquetas para los objetos de la configuración.

Los valores de *CITY* oscilan entre 1 y 12. A estos números enteros se les asignan valores, donde:

- 1 = ATLANTA
- 2 = BOSTON
- 3 = CINCINATI
- 4 = COLÓN
- 5 = DALLAS
- 6 = INDNPLIS
- 7 = LITTROCK
- 8 = LOSÁNGELES
- 9 = MEMPHIS
- 10 = San Luis
- 11 = SPOKANE
- 12 = TAMPA

PROXMAP utiliza estas etiquetas de valor para identificar objetos en tablas y representaciones gráficas, garantizando que se muestren los nombres de los objetos en lugar de códigos numéricos.

Configuración del análisis PROXMAP

Ir a **«** > **Anal** > **izar» > «Mapeo de proximidad»**

Las variables de la matriz de proximidad se especifican como una única fuente de proximidad en el cuadro de diálogo «Mapeo de proximidad». PROXMAP admite datos de proximidad almacenados tanto en forma de matriz triangular inferior (incluida la diagonal) como en forma de matriz simétrica completa, siempre que la matriz esté correctamente especificada.

Pestaña «Datos» (Especificación de la matriz de proximidad y las etiquetas de los objetos)
Define la fuente de proximidad:

1. Selecciona las variables *desde ATLANTA* hasta *TAMPA* en la lista de variables y arrástralas a **la sección «Variables» en «Fuente** ».
2. Introduzca un nombre en el campo «**Etiqueta de origen** » (por ejemplo, *«Distancias de vuelo entre ciudades»*).
3. Haz clic en **«Añadir fuente»** para crear una fuente (por ejemplo, *$SRC001*).
4. En «**Tipo de datos** », selecciona «**Datos de proximidad** ».
5. Mueve la variable *CITY* a **las etiquetas de objetos**

El **número actual de objetos = 12** refleja el número de objetos definidos por la matriz de proximidad y confirma que habrá doce ciudades representadas en la configuración.

Pestaña «Modelo» (Especificación del modelo PROXMAP)
El análisis se basa en una única fuente de proximidad ($SRC001), por lo que debe seleccionar las siguientes opciones:

1. En **el «Modelo de tres vías»**, selecciona la opción «**Identidad** ».

2. En **«Condicionalidad/Transformación»**, selecciona **«Sin condiciones»**
3. Haz clic en **«Definir transformación»** y selecciona las opciones **«Ordinal»** y «**Permitir desempatar** ».

   Esto indica que las proximidades se tratan como diferencias ordinales. Se mantiene el orden de las proximidades, y los valores iguales en los datos siguen estando empatados en la transformación monótona.

   Si es necesario, puedes modificar el tipo de transformación (por ejemplo, lineal, exponencial, spline, etc.).

El número actual de dimensiones = 2 indica que se estimará una configuración bidimensional.

Esta configuración se ajusta a las especificaciones utilizadas para la comparación con ALSCAL.

Pestaña «Criterios» (Especificación de la dimensionalidad, la convergencia y la estandarización)

1. **Dimensionalidad** : el **número de dimensiones** está establecido en *2*

   Esto indica que se estimará una configuración bidimensional.
2. **Convergencia del algoritmo** : Los criterios de convergencia controlan el proceso de optimización iterativo. Se especifican los siguientes ajustes:
   - **Función mínima:** 0.000001
   - **Diferencia entre funciones:** 0.000000000001
   - **Diferencias de configuración:** 0.000001
   - **Número máximo de iteraciones:** 1000

   Estos valores determinan cuándo se detiene el algoritmo iterativo. La convergencia se alcanza cuando las mejoras en la función de pérdida y los cambios en la configuración se sitúan por debajo de los límites de tolerancia especificados, o cuando se alcanza el número máximo de iteraciones.
3. **Normalización** : Seleccionar «**Suma de cuadrados = N** ». Esta opción resulta especialmente relevante para la estandarización de datos multivariantes.
4. **Dimensión** : Seleccionar **por dimensión**. Con esta opción puedes ver el reflejo de cada una de las dimensiones. Marca la casilla situada debajo de **«Reflection»** en la dimensión 2.

   Las reflexiones solo afectan a la orientación (signo) de los ejes y no modifican las distancias ni el ajuste del modelo. Normalmente se utilizan para ajustar la solución a una configuración de referencia (por ejemplo, al compararla con una solución de ALSCAL).

Pestaña «Resultados»: subpestaña «Modelo» (Solicitud de tablas y gráficos)

1. En la sección «Tablas», seleccione **«Coordenadas»** para que se genere en la salida una tabla que contenga las coordenadas de los objetos en el espacio común.

   Las opciones **«Direcciones variables»** y **«Ponderaciones espaciales»** no están seleccionadas, ya que este análisis se basa en una única fuente de proximidad y no incluye fuentes adicionales ni variables multivariantes.
2. En «Gráficos», dentro de la sección «Espacio común», selecciona «**Configuración** ».

   Esto genera un gráfico de la configuración del objeto para la fuente $SRC001 (por ejemplo, las distancias de vuelo entre ciudades), con etiquetas que se indican en la variable especificada en la pestaña **«Datos»** (por ejemplo, *CITY*).
3. **Salida adicional** : En la lista **«Fuentes»**, desplaza « *$SRC001* » a **«Transformación»**, «**Ajuste** » y «**Shepard** ».

   Se generan los siguientes gráficos:
   - Un gráfico que muestra **las proximidades transformadas** en el eje vertical frente a las proximidades originales en el eje horizontal.
   - Un diagrama de dispersión que muestra las **distancias ajustadas** en la configuración frente a las **proximidades transformadas**.

- Un gráfico de dispersión de diagnóstico que muestra las distancias ajustadas, junto con las proximidades transformadas frente a las proximidades originales, para visualizar la bondad de ajuste en el escalado multidimensional no métrico (mapeo de proximidades).

Se solicita una solución bidimensional, acorde con la dimensionalidad utilizada en el análisis ALSCAL.

Sintaxis del mandato

Además, consulta la siguiente sintaxis de comando utilizada para el análisis mediante ALSCAL y PROXMAP.

Sintaxis de los comandos utilizados para el análisis con ALSCAL

```
ALSCAL
  VARIABLES=ATLANTA BOSTON CINCNATI COLUMBUS DALLAS INDNPLIS LITTROCK LOSANGEL MEMPHIS
STLOUIS
    SPOKANE TAMPA
  /SHAPE=SYMMETRIC
  /LEVEL=ORDINAL
  /CONDITION=MATRIX
  /MODEL=EUCLID
  /CRITERIA=CONVERGE(0.001) STRESSMIN(0.005) ITER(30) CUTOFF(0) DIMENS(2,2)
  /PLOT.
```

Sintaxis de los comandos utilizados para el análisis con PROXMAP

```
PROXMAP
  /SOURCEID NAME=$SRC001 LABEL='Cities flying distances'  DATA=ATLANTA BOSTON CINCNATI
COLUMBUS DALLAS INDNPLIS
    LITTROCK LOSANGEL MEMPHIS STLOUIS SPOKANE TAMPA
  /DATA PROXIMITIES
  /OBJECTID LABELS=CITY
  /CONDITION UNCONDITIONAL
  /TRANSFORMATIONS SOURCES=$SRC001 FUNCTION=ORDINAL APPROACH=KEEPTIES
  /MODEL IDENTITY
  /CRITERIA DIMENSIONS=2 MAXIMUM=1000 MINIMUM=0.000001 RELATIVE=0.000000001
ABSOLUTE=0.000001
    STANDARDIZATION='N' REFLECTION=DIMENSIONS NUMBERS=1 -1
  /PRINT DATA(RAW) MODEL (COORDINATES) CRITERIA (SUMMARY ITERATIONS)
  /PLOT MODEL (CONFIGURATION TRANSFORMATION ($SRC001) FIT ($SRC001) SHEPARD ($SRC001)).
```

Gráficos de resultados y de diagnóstico

El resultado incluye una tabla con las coordenadas de los objetos en el espacio común de objetos y un gráfico de configuración que muestra la representación espacial de las ciudades.

Se solicitan gráficos de diagnóstico adicionales, entre los que se incluyen gráficos de transformación, gráficos de ajuste y gráficos de Shepard. Estos gráficos muestran las relaciones entre las proximidades originales, las proximidades transformadas y las distancias ajustadas, y ofrecen información visual sobre la calidad del ajuste de la solución.

La terminología de PROXMAP difiere de la de ALSCAL. Por ejemplo, los estímulos de ALSCAL se denominan «objetos» en PROXMAP, y las disparidades de ALSCAL se denominan «proximidades transformadas». El gráfico de Shepard en PROXMAP combina la información que se muestra en los gráficos de ajuste no lineal y de transformación de ALSCAL.

Resultados e interpretación

Cuando se ejecuta PROXMAP con parámetros similares a los de ALSCAL, las configuraciones resultantes son muy similares. Las diferencias entre las soluciones se limitan a un factor de escala general y a posibles reflejos en los ejes.

Los valores de las coordenadas pueden diferir numéricamente, pero representan la misma configuración subyacente, salvo por la escala. PROXMAP indica las coordenadas basándose en la escala original de las proximidades.

Una comparación visual de los gráficos de configuración muestra que se conservan las posiciones relativas de los objetos, lo que confirma que PROXMAP reproduce correctamente la solución de ALSCAL.

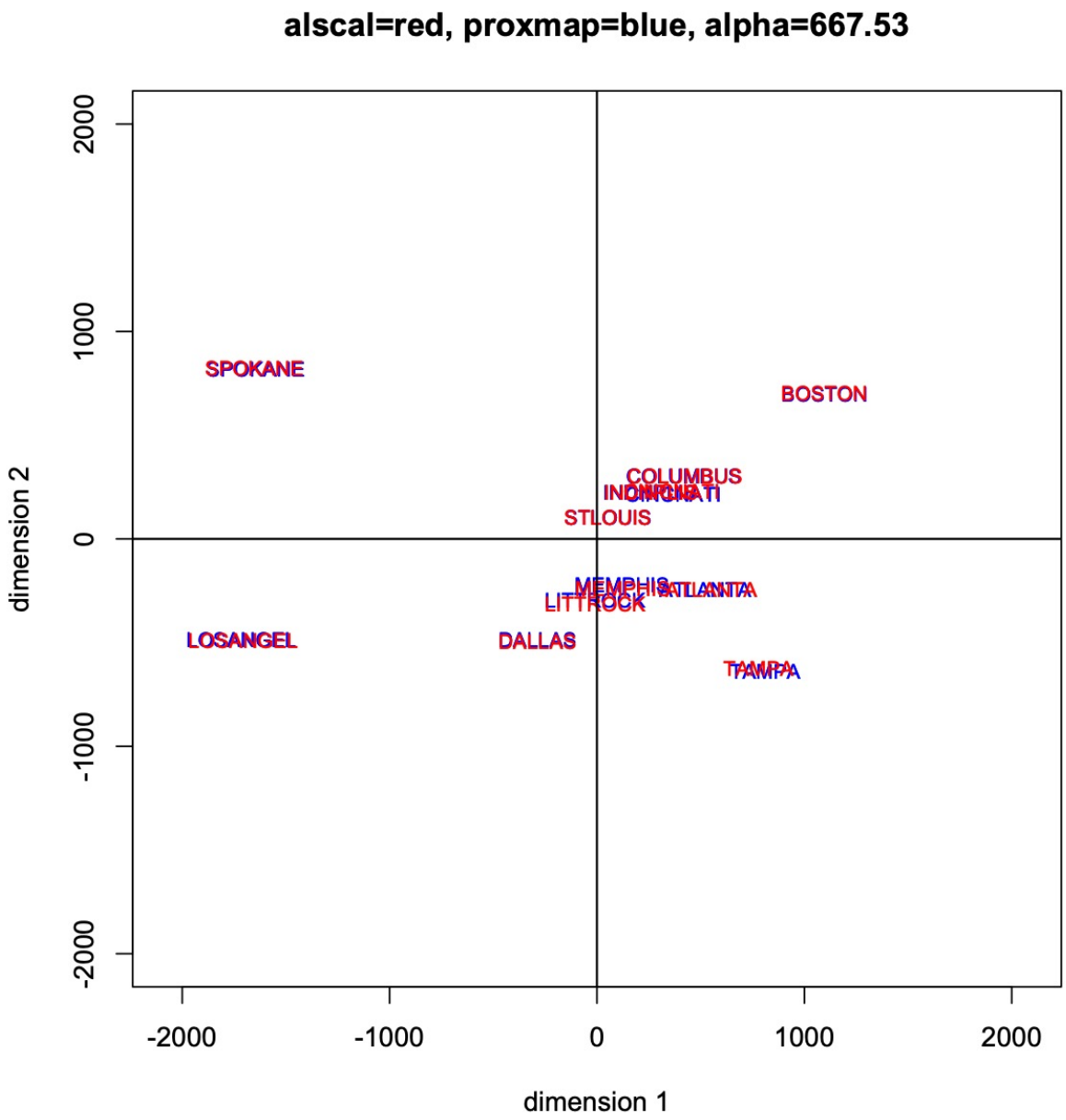

Figura 3. Configuración de estímulos de ALSCAL: coordenadas PROXMAP del espacio común

Puntos clave para los usuarios de ALSCAL

- Los datos de proximidad existentes para ALSCAL pueden reutilizarse en PROXMAP, incluidos los datos almacenados en forma de matriz triangular inferior.
- Se necesita una variable de etiqueta de objeto adicional para identificar los objetos en los resultados y gráficos de PROXMAP.
- Al comparar los resultados de PROXMAP con los de ALSCAL, pueden producirse diferencias en la orientación de las coordenadas. Los reflejos solo afectan a la orientación de los ejes y no modifican las distancias ni la calidad del ajuste.
- PROXMAP utiliza terminología actualizada y métodos modernos de escalamiento multidimensional, al tiempo que mantiene la continuidad conceptual con ALSCAL.

Estadísticos de la razón

El procedimiento Estadísticos de la razón proporciona una amplia lista de estadísticos de resumen para describir la razón entre dos variables de escala.

Se pueden ordenar los resultados por los valores de una variable de agrupación, en orden ascendente o descendente. Se puede eliminar de los resultados el informe de los estadísticos de la razón y almacenar los resultados en un archivo externo.

Ejemplo

¿Existe una buena uniformidad en la razón entre el precio de tasación y el precio de venta de viviendas en cada una de las cinco regiones? En los resultados, se puede descubrir que la distribución de las razones varía considerablemente entre regiones.

Estadísticos

Mediana, media, media ponderada, intervalos de confianza, coeficiente de dispersión (COD), coeficiente de variación (COV), coeficiente de variación centrado en la mediana, coeficiente de variación centrado en la media, sesgo relacionado con el precio (PRB), diferencial de precios (PRD), desviación estándar, desviación absoluta media (AAD), rango, valores mínimos y máximos, y el índice de concentración calculado para un rango o porcentaje especificado por el usuario dentro de la razón mediana.

Consideraciones sobre los datos

Datos

Utilice códigos numéricos o cadenas para codificar las variables de agrupación (mediciones de nivel nominal u ordinal).

Supuestos

Las variables que definen el numerador y el denominador de la razón deben ser variables de escala, que toman valores positivos.

Obtención de estadísticos de la razón

1. En los menús seleccione:

   Analizar > **Estadísticos descriptivos** > **Razón...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una variable de numerador.
3. Seleccione una variable de denominador.

Si lo desea:

- Seleccione una variable de agrupación y especificar el orden de los grupos en los resultados.
- Elija si desea mostrar los resultados en el Visor.
- Elija si desea añadir una palabra clave **N** al subcomando **PRINT**. El valor añade los tamaños de muestra a la tabla de estadísticos de los resultados.
- Elija si desea guardar los resultados en un archivo externo para un uso posterior y especificar el nombre del archivo en el que se van a guardar los resultados.

Estadísticos de la razón: Estadísticos

Tendencia central

Las medidas de tendencia central son estadísticos que describen la distribución de las razones.

Mediana

Un valor tal que el número de razones menores que este valor es igual al número de razones mayores que el mismo.

Media
El resultado de sumar las razones y dividir la suma entre el número total de razones.

Media ponderada
El resultado de dividir la media del numerador por la media del denominador. La media ponderada es también la media de las razones ponderadas por el denominador.

Intervalos de confianza para PRB y medidas de tendencia central
Muestra los intervalos de confianza para PRB, la media, la mediana y la media ponderada (si se solicita). Especifique un valor mayor o igual que 0 y menor que 100 como nivel de confianza. El valor predeterminado es 95%.

Dispersión
Estos estadísticos miden la cantidad de variación, o dispersión, en los valores observados.

DAP
La desviación absoluta promedio es el resultado de sumar las desviaciones absolutas de las razones respecto a la mediana y dividir el resultado entre el número total de razones.

CDD
El coeficiente de dispersión es el resultado de expresar la desviación absoluta promedio como un porcentaje de la mediana.

CV
CV es el coeficiente de variación. Es la relación entre la desviación típica y la media. Los valores se expresan en porcentajes.

PRB
El sesgo relativo al precio. Un índice que indica si la evaluación de las razones de precios son sistemáticamente más altas o más bajas para propiedades con precios más altos. Realiza la regresión de las diferencias de porcentaje en las razones de evaluación a partir de la razón de la mediana en los logaritmos en base 2 de las medidas de proxy de valor calculadas como los promedios de los precios de venta y las razones de valores evaluados a la razón de la mediana. Proporciona un cambio de porcentaje en razones de evaluación para un cambio del 100 por ciento en el valor.

DRP
El diferencial relativo al precio, también conocido como el índice de regresibilidad, es el resultado de dividir la media por la media ponderada.

Desviación estándar
La desviación estándar es el resultado de sumar las desviaciones cuadráticas de las razones respecto a la media, dividir la suma por el número total de razones menos uno y extraer la raíz cuadrada positiva.

Rango
El rango es el resultado de restar la razón mínima de la razón máxima.

Mínimo
El mínimo es la razón mínima.

Máximo
El máximo es la razón mayor.

Índice de concentración
El coeficiente de concentración mide el porcentaje de razones que están dentro de un intervalo. Se puede calcular de dos maneras:

Entre proporciones
En este caso, el intervalo se define de forma explícita especificando los valores superior e inferior del intervalo. Introduzca valores para las proporciones superior e inferior y pulse en **Añadir** para obtener un intervalo.

Dentro del porcentaje de la mediana
En este caso, el intervalo se define de forma implícita al especificar el porcentaje de la mediana. Introduzca un valor entre 0 y 100 y pulse **Añadir**. El límite inferior del intervalo será igual a (1 – 0.01 × valor) × mediana, y el límite superior será igual a (1 + 0.01 × valor) × mediana.

Gráficos P-P

El procedimiento de gráficos P-P produce gráficos de probabilidad de una o varias secuencias o variables de serie temporal. Las variables se pueden estandarizar, diferenciar o transformar antes de representarlas en gráficos.

Entre las distribuciones de contraste disponibles se incluyen Beta, Chi-Cuadrado, Exponencial, Gamma, Semi-Normal, Laplace, Logística, Lognormal, Normal, Pareto, t de Student, Weibull y Uniforme. Según la distribución seleccionada, podrá especificar los grados de libertad y otros parámetros.

- Se obtendrán gráficos de probabilidad para los valores transformados. Entre las opciones de transformación se incluyen Transformación log natural, Estandarizar los valores, Diferenciar y Diferenciar ciclo.
- Puede especificar los métodos para calcular las distribuciones esperadas y para deshacer los "empates" (varias observaciones con el mismo valor).

Distribución de prueba

Especifique un tipo de distribución para los datos. La lista desplegable proporciona las opciones siguientes:

Beta

Distribución Beta. Los parámetros forma1 y format2 *a* y *b* deben ser positivos. Si no se especifican, DISTRIBUTION los estima de la media de la muestra y de la desviación estándar de ejemplo. Todas las observaciones deben estar en el rango 0-1, ambos inclusive.

Chi-cuadrado

Distribución de chi-cuadrado. Debe especificar los grados de libertad (*df*). No se permiten observaciones negativas.

Exponencial

Distribución exponencial. El parámetro de escala *a* debe ser positivo. Si no se especifica el parámetro, DISTRIBUTION lo estima de la media de la muestra. No se permiten observaciones negativas.

Gamma

Distribución gamma. Los parámetros de forma y escala *a* y *b* deben ser positivos. Si no se especifican, DISTRIBUTION los estima de la media de la muestra y de la desviación estándar de ejemplo. No se permiten observaciones negativas.

Semi-normal

Distribución semi-normal. Se supone que los datos están libres de ubicación o están centralizados. (Parámetro de ubicación=0.) Puede especificar el parámetro de escala *a* o permitir que DISTRIBUTION lo calcule utilizando el método de máxima verosimilitud.

Laplace

Distribución Laplace o exponencial doble. LAPLACE toma una ubicación y un parámetro de escala (*a* y *b*). El parámetro de escala (*b*) debe ser positivo. Si no se especifican los parámetros, DISTRIBUTION los estima de la media de la muestra y de la desviación estándar de ejemplo.

Logística

Distribución logística. LOGISTIC toma una ubicación y un parámetro de escala (*a* y *b*). El parámetro de escala (*b*) debe ser positivo. Si no se especifican los parámetros, DISTRIBUTION los estima de la media de la muestra y de la desviación estándar de ejemplo.

Lognormal

Distribución Lognormal. Los parámetros de escala y forma *a* y *b* deben ser positivos. Si no se especifican, DISTRIBUTION los estima de la media y la desviación estándar del logaritmo natural de los datos de la muestra. No se permiten observaciones negativas.

Normal

Distribución normal. El parámetro de ubicación *a* puede ser cualquier valor numérico, mientras que el parámetro de escala *b* debe ser positivo. Si no se especifican, DISTRIBUTION los estima de la media de la muestra y de la desviación estándar de ejemplo.

Pareto
Distribución Pareto. Los parámetros de umbral y forma *a* y *b* deben ser positivos. Si no se especifican, DISTRIBUTION asume que *a* es igual a la mínima observación y calcula *b* por el método de máxima verosimilitud. No se permiten observaciones negativas.

T de Student
Distribución t *Student's.* Debe especificar los grados de libertad (*df).*

Uniforme
Distribución uniforme. UNIFORM toma un parámetro mínimo y máximo (*a* y *b).* El parámetro *a* debe ser igual o mayor que *b.* Si no se especifican los parámetros, DISTRIBUTION los presupone a partir de los datos de ejemplo.

Weibull
Distribución de Weibull. Los parámetros de escala y forma *a* y *b* deben ser positivos. Si no se especifican, DISTRIBUTION los calcula utilizando el método de mínimo cuadrado. No se permiten observaciones negativas.

Parámetros de distribución
Proporciona opciones de estrategia de distribución y de parámetros.

Estimar a partir de los datos
Cuando se selecciona, este valor estima los parámetros de distribución basados en los datos y el tipo de distribución seleccionado.

Especifique
Cuando se selecciona, puede especificar los parámetros de distribución para el tipo de distribución seleccionado.

Nota: Los parámetros disponibles variarán según el tipo de distribución seleccionado.

Transformar
Las opciones proporcionadas establecen los valores de transformación y periodicidad.

Transformación log natural
Transforma los datos utilizando el algoritmo natural (base *e*) para eliminar la amplitud variable.

Estandarizar valores
Transforma la secuencia de variables de serie temporal en una muestra con una media de 0 y una desviación estándar de 1.

Diferencia
Especifica el grado de diferenciación que se utiliza antes de representar para convertir una variable no estacionaria en una variable estacionaria con una media y varianza constantes. Especifique un valor adecuado en el campo.

Diferencia estacional
Si la variable muestra un patrón estacional o periódico, puede utilizar este ajuste para desestacionalizar la variable antes de realizar la representación.

Nota: Este valor sólo se habilita cuando se selecciona una variable de secuencia o de serie temporal, con una periodicidad definida, como una de las variables cuantitativas.

Fórmula de estimación de proporción
Las opciones proporcionadas establecen la fórmula que se utiliza para estimar las proporciones.

De Blom
Crea una nueva variable de clasificación basada en estimaciones de proporción que utiliza la fórmula `(r-3/8) / (w+1/4)`, donde `w` es la suma de los pesos de los casos y `r` es el rango.

Rankit
Utiliza la fórmula `(r-1/2) / w`, donde `w` es el número de observaciones y `r` es el rango, que va de 1 a `w`.

De Tukey
Utiliza la fórmula `(r-1/3) / (w+1/3)`, donde `r` es el rango y `w` es la suma de las ponderaciones de los casos.

De Van der Waerden
La transformación de Van der Waerden, definida por la fórmula `r/(w+1)`, donde `w` es la suma de los pesos de los casos y `r` es el rango, que va de 1 a `w`.

Rango asignado a empates
Las opciones proporcionadas controlan el método para determinar cómo manejar los valores de empate. La tabla siguiente muestra cómo los distintos métodos asignan rangos a los valores empatados.

Tabla 6. Métodos de clasificación y resultados

Valor	Media	Bajo	Alto	Romper los empates arbitrariamente
10	1	1	1	1
19	3	2	4	2
19	3	2	4	2
19	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Obtención de gráficos de probabilidad P-P

1. En los menús seleccione:

   Analizar > **Estadísticas descriptivas** > **Gráficos P-P...**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione una o más variables numéricas y desplácelas al campo Variables.
3. Seleccione una distribución de prueba.

Si lo desea, puede seleccionar opciones de transformación para obtener gráficos de probabilidad para los valores transformados y especificar un método para calcular las distribuciones esperadas.

Gráficos Q-Q

El procedimiento de gráficos Q-Q genera gráficos de probabilidad para valores transformados. Entre las distribuciones de contraste disponibles se incluyen Beta, Chi-Cuadrado, Exponencial, Gamma, Semi-Normal, Laplace, Logística, Lognormal, Normal, Pareto, t de Student, Weibull y Uniforme. Según la distribución seleccionada, podrá especificar los grados de libertad y otros parámetros.

- Se obtendrán gráficos de probabilidad para los valores transformados. Entre las opciones de transformación se incluyen Transformación log natural, Estandarizar los valores, Diferenciar y Diferenciar ciclo.
- Puede especificar los métodos para calcular las distribuciones esperadas y para deshacer los "empates" (varias observaciones con el mismo valor).

Distribución de prueba
Especifique un tipo de distribución para los datos. La lista desplegable proporciona las opciones siguientes:

Beta
Distribución Beta. Los parámetros forma1 y format2 *a* y *b* deben ser positivos. Si no se especifican, `DISTRIBUTION` los estima de la media de la muestra y de la desviación estándar de ejemplo. Todas las observaciones deben estar en el rango 0-1, ambos inclusive.

Chi-cuadrado
Distribución de chi-cuadrado. Debe especificar los grados de libertad (*df)*. No se permiten observaciones negativas.

Exponencial
Distribución exponencial. El parámetro de escala *a* debe ser positivo. Si no se especifica el parámetro, DISTRIBUTION lo estima de la media de la muestra. No se permiten observaciones negativas.

Gamma
Distribución gamma. Los parámetros de forma y escala *a* y *b* deben ser positivos. Si no se especifican, DISTRIBUTION los estima de la media de la muestra y de la desviación estándar de ejemplo. No se permiten observaciones negativas.

Semi-normal
Distribución semi-normal. Se supone que los datos están libres de ubicación o están centralizados. (Parámetro de ubicación=0.) Puede especificar el parámetro de escala *a* o permitir que DISTRIBUTION lo calcule utilizando el método de máxima verosimilitud.

Laplace
Distribución Laplace o exponencial doble. LAPLACE toma una ubicación y un parámetro de escala (*a* y *b).* El parámetro de escala *(b*) debe ser positivo. Si no se especifican los parámetros, DISTRIBUTION los estima de la media de la muestra y de la desviación estándar de ejemplo.

Logística
Distribución logística. LOGISTIC toma una ubicación y un parámetro de escala (*a* y *b).* El parámetro de escala *(b*) debe ser positivo. Si no se especifican los parámetros, DISTRIBUTION los estima de la media de la muestra y de la desviación estándar de ejemplo.

Lognormal
Distribución Lognormal. Los parámetros de escala y forma *a* y *b* deben ser positivos. Si no se especifican, DISTRIBUTION los estima de la media y la desviación estándar del logaritmo natural de los datos de la muestra. No se permiten observaciones negativas.

Normal
Distribución normal. El parámetro de ubicación *a* puede ser cualquier valor numérico, mientras que el parámetro de escala *b* debe ser positivo. Si no se especifican, DISTRIBUTION los estima de la media de la muestra y de la desviación estándar de ejemplo.

Pareto
Distribución Pareto. Los parámetros de umbral y forma *a* y *b* deben ser positivos. Si no se especifican, DISTRIBUTION asume que *a* es igual a la mínima observación y calcula *b* por el método de máxima verosimilitud. No se permiten observaciones negativas.

T de Student
Distribución t *Student's.* Debe especificar los grados de libertad (*df).*

Uniforme
Distribución uniforme. UNIFORM toma un parámetro mínimo y máximo (*a* y *b).* El parámetro *a* debe ser igual o mayor que *b.* Si no se especifican los parámetros, DISTRIBUTION los presupone a partir de los datos de ejemplo.

Weibull
Distribución de Weibull. Los parámetros de escala y forma *a* y *b* deben ser positivos. Si no se especifican, DISTRIBUTION los calcula utilizando el método de mínimo cuadrado. No se permiten observaciones negativas.

Parámetros de distribución
Proporciona opciones de estrategia de distribución y de parámetros.

Estimar a partir de los datos
Cuando se selecciona, este valor estima los parámetros de distribución basados en los datos y el tipo de distribución seleccionado.

Especifique
Cuando se selecciona, puede especificar los parámetros de distribución para el tipo de distribución seleccionado.

Nota: Los parámetros disponibles variarán según el tipo de distribución seleccionado.

Transformar
Las opciones proporcionadas establecen los valores de transformación y periodicidad.

Transformación log natural
Transforma los datos utilizando el algoritmo natural (base *e*) para eliminar la amplitud variable.

Estandarizar valores
Transforma la secuencia de variables de serie temporal en una muestra con una media de 0 y una desviación estándar de 1.

Diferencia
Especifica el grado de diferenciación que se utiliza antes de representar para convertir una variable no estacionaria en una variable estacionaria con una media y varianza constantes. Especifique un valor adecuado en el campo.

Diferencia estacional
Si la variable muestra un patrón estacional o periódico, puede utilizar este ajuste para desestacionalizar la variable antes de realizar la representación.

Nota: Este valor sólo se habilita cuando se selecciona una variable de secuencia o de serie temporal, con una periodicidad definida, como una de las variables cuantitativas.

Fórmula de estimación de proporción
Las opciones proporcionadas establecen la fórmula que se utiliza para estimar las proporciones.

De Blom
Crea una nueva variable de clasificación basada en estimaciones de proporción que utiliza la fórmula `(r-3/8) / (w+1/4)`, donde `w` es la suma de los pesos de los casos y `r` es el rango.

Rankit
Utiliza la fórmula `(r-1/2) / w`, donde `w` es el número de observaciones y `r` es el rango, que va de 1 a `w`.

De Tukey
Utiliza la fórmula `(r-1/3) / (w+1/3)`, donde `r` es el rango y `w` es la suma de las ponderaciones de los casos.

De Van der Waerden
La transformación de Van der Waerden, definida por la fórmula `r/(w+1)`, donde `w` es la suma de los pesos de los casos y `r` es el rango, que va de 1 a `w`.

Rango asignado a empates
Las opciones proporcionadas controlan el método para determinar cómo manejar los valores de empate. La tabla siguiente muestra cómo los distintos métodos asignan rangos a los valores empatados.

Tabla 7. Métodos de clasificación y resultados

Valor	Media	Bajo	Alto	Romper los empates arbitrariamente
10	1	1	1	1
19	3	2	4	2
19	3	2	4	2
19	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Obtención de gráficos de probabilidad Q-Q

1. En los menús seleccione:

Analizar > **Estadísticas descriptivas** > **Gráficos Q-Q...**

Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.

2. Seleccione una o más variables numéricas y desplácelas al campo Variables.
3. Seleccione una distribución de prueba.

Si lo desea, puede seleccionar opciones de transformación para obtener gráficos de probabilidad para los valores transformados y especificar un método para calcular las distribuciones esperadas.

Análisis ROC

El Análisis ROC (Receiver operating characteristic o característica operativa del receptor) es una forma útil de evaluar la precisión de las predicciones de modelo al trazar la sensibilidad frente a la especificidad de una prueba de clasificación (ya que el umbral varía en todo un rango de resultados de pruebas de diagnostico). El área completa bajo una curva ROC determinada, o AUC, formula una estadística importante que representa la probabilidad de que la predicción esté en el orden correcto cuando se observa una variable de prueba (para un sujeto que se selecciona aleatoriamente del grupo de casos, y el otro se selecciona aleatoriamente del grupo de control). El Análisis ROC admite la inferencia en relación con una sola AUC, curvas de precisión/exhaustividad (PR), y proporciona opciones para comparar dos curvas ROC que se generan de grupos independientes o sujetos emparejados.

El antiguo procedimiento de Curva ROC soporta la inferencia estadística sobre una sola curva ROC. Esto también puede ser recuperado por el nuevo procedimiento de análisis ROC. Además, el nuevo procedimiento de análisis ROC puede comparar dos curvas ROC generadas a partir de grupos independientes o sujetos emparejados.

Las curvas PR representan la precisión frente a la exhaustividad, tienden a ser más informativas cuando las muestras de datos observados son muy asimétricas, y proporcionan una alternativa a las curvas ROC para datos con una alta asimetría en la distribución de clase.

Ejemplo
Un banco tiene interés en clasificar a sus clientes dependiendo de si se retrasarán o no en el pago de sus préstamos; por tanto, se desarrollan modelos especiales para tomar estas decisiones. Se puede utilizar el Análisis ROC para evaluar y valorar la precisión de las predicciones del modelo.

Estadísticos
AUC, grupo negativo, valores perdidos, clasificación positiva, valor de corte, grado de convicción, intervalo de confianza asintótico bilateral, distribución, error estándar, diseño de grupo independiente, diseño de muestra emparejada, supuesto no paramétrico, supuesto de distribución exponencial binegativo, punto medio, punto de corte, curva PR, interpolación por pasos, significación asintótica (bilateral), Sensibilidad y (1 - Especificidad), Precisión y Exhaustividad.

Métodos
Se comparan las áreas bajo dos curvas ROC, que se generan a partir de grupos independientes o sujetos emparejados. Comparar dos curvas ROC puede proporcionar más información en cuanto a la precisión que resulta de dos acercamientos de diagnósticos comparativos.

Consideraciones de datos sobre el Análisis ROC

Datos
Las curvas PR representan la precisión frente a la exhaustividad y tienden a ser más informativas cuando las muestras de datos observados son muy asimétricas. Una interpolación lineal simple puede producir erróneamente una estimación excesivamente optimista de una curva PR.

Supuestos
La predicción estará en el orden correcto cuando se observa una variable de prueba para un sujeto seleccionado aleatoriamente del grupo de casos, y el otro seleccionado aleatoriamente del grupo de control. Cada grupo definido contendrá al menos una observación válida. Sólo se utiliza una variable de agrupación única para un procedimiento único.

Obtención de un análisis ROC

1. En los menús seleccione:

   Analizar > **Clasificar** > **Análisis ROC**

   Nota: Los campos resaltados en rojo son obligatorios. Los botones **Pegar** y **Aceptar** se activan después de introducir valores válidos en todos los campos obligatorios.
2. Seleccione uno o más **Variable de prueba** variables de probabilidad de prueba.
3. Seleccione uno **Variable de estado** .
4. Identifica el *positivo* **Valor de la variable de estado** .
5. Si lo desea, seleccione la opción **Diseño de muestras emparejadas** o seleccione una única variable de agrupación (no puede seleccionar ambas opciones).
   - Utilice el valor **Diseño de muestras emparejadas** para solicitar el diseño de muestras emparejadas para las variables de prueba. El diseño de muestras emparejadas compara dos curvas ROC en un escenario de muestras emparejadas cuando se miden varios valores de prueba en los mismos sujetos que están asociados con una variable de estado.

     Nota: Si se selecciona **Diseño de muestras emparejadas**, se inhabilitan las opciones **Variable de agrupación** y **Asunción de distribución** (en el diálogo Opciones).
   - Cuando se selecciona una variable de agrupación numérica, puede pulsar **Definir grupos ...** solicitar el diseño de grupo independiente para las variables de prueba y especificar dos valores, un punto medio o un punto de corte.
6. Si lo desea, pulse Opciones para definir los valores de clasificación, dirección de la prueba, parámetros de error estándar y valores perdidos.
7. Si lo desea, pulse Visualización para definir los valores de trazado e impresión (que incluyen los valores de Curva ROC, Curva de precisión/exhaustividad y Calidad de modelo).
8. Pulse **Aceptar**.

Análisis ROC: Opciones

Puede seleccionar las opciones siguientes para su análisis ROC:

Clasificación
Permite especificar si se debe incluir o excluir el valor del punto de corte al realizar una clasificación *positiva*. Este ajuste no afecta a los resultados.

Dirección de la prueba
Proporciona opciones para especificar qué dirección de la variable de resultado de la prueba indica un creciente grado de convicción de que el sujeto dará positivo en la prueba.

Parámetros para el error estándar del área
Permite especificar el método de estimación del error estándar del área situada bajo la curva. Los métodos disponibles son el no paramétrico y el exponencial binegativo. El valor predeterminado **No paramétrico** proporciona estimaciones bajo el supuesto no paramétrico. El valor **Exponencial binegativo** proporciona estimaciones bajo el supuesto de distribución exponencial binegativo.

La sección también le permite especificar el nivel de confianza del intervalo de confianza asintótico bilateral de AUC. El rango disponible es 0,0% a 100,0% (el valor predeterminado es 95%).

Nota: El valor solo se aplica al diseño de grupo independiente y no tiene efecto en el diseño de muestras emparejadas.

Valores perdidos
Permite especificar el tratamiento que reciben los valores perdidos. Cuando el valor no está seleccionado, se excluyen tanto los valores perdidos del usuario como los valores perdidos del sistema. Cuando el valor está seleccionado, se tratan como válidos los valores perdidos del usuario y se excluyen los valores perdidos del sistema. Los casos con valores perdidos del sistema, ya sea en la variable de prueba o en la variable de estado, se excluyen siempre del análisis.

Análisis ROC: Visualización

Puede especificar los valores de visualización siguientes para el análisis ROC:

Gráfico
Proporciona opciones para trazar las curvas ROC y precisión/exhaustividad.

Curva ROC
Cuando se selecciona, se muestra un gráfico Curva ROC en la salida. Seleccione **Con línea diagonal de referencia** para trazar una línea diagonal de referencia en el gráfico Curva ROC.

Curva de precisión/exhaustividad
Cuando se selecciona, se muestra en la salida un gráfico Curva de precisión/exhaustividad. Las curvas de precisión/exhaustividad tienden a ser más informativas cuando las muestras de datos observados son muy asimétricas, y proporcionan una alternativa a las curvas ROC para datos con una alta asimetría en la distribución de clase. El valor predeterminado **Interpolar en los verdaderos positivos** realiza la interpolación por pasos en los verdaderos positivos. El valor **Interpolar en los falsos positivos** realiza la interpolación por pasos en los falsos positivos.

Calidad global de modelo
El valor controla si se crea un gráfico de barras o no para mostrar el valor del límite inferior del intervalo de confianza del AUC estimado. De forma predeterminada, el valor no está seleccionado, lo cual suprime el gráfico de barras.

Imprimir
Proporciona opciones para definir la salida de las estadísticas correspondientes.

Error estándar e intervalo de confianza
El valor controla qué estadísticos se muestran en la tabla "Área bajo la curva". Cuando el valor no está seleccionado, sólo se visualiza el valor estimado de AUC. Cuando se selecciona el valor, se muestran estadísticos adicionales, incluidos el error estándar de AUC, la significación asintótica (bilateral) y los límites de intervalo de confianza asintótica bajo la hipótesis nula.

Puntos de coordenadas de curva ROC
El valor controla los puntos de coordenadas de la Curva ROC, junto con los valores de corte. Cuando el valor no está seleccionado, se suprime la salida de los puntos de coordenadas. Cuando se selecciona el valor, se dan los pares de valores de Sensibilidad y (1 - Especificidad) con los valores de corte de cada curva ROC.

Índice de Youden

Muestra el valor del índice de Youden para cada valor de corte de la curva ROC.

Puntos de coordenadas de la curva de precisión/exhaustividad
El valor controla los puntos de coordenadas de la curva de precisión/exhaustividad, junto con los valores de corte. Cuando el valor no está seleccionado, se suprime la salida de los puntos de coordenadas. Cuando se selecciona el valor, se dan los pares de valores de precisión y exhaustividad con los valores de corte de cada curva de precisión/exhaustividad.

Métricas de evaluación de clasificador
Este parámetro controla la visualización de la tabla Métricas de evaluación de clasificador en el resultado. La tabla muestra lo bien que un modelo de clasificación se ajusta a los datos en comparación con una asignación aleatoria y proporciona la información siguiente:

- Las variables de prueba especificadas por el usuario
- Información de grupo
- Índice de Gini (el índice de Gini es 2*AUC - 1, donde AUC es el área debajo de la curva ROC)
- Valores de K-S máximo y de Corte

Análisis ROC: Definir grupos (serie)

Para las variables de agrupación de series, especifique una serie para el grupo 1 y otro valor para el grupo 2, como por ejemplo *Sí* y *no*. Los casos con otras series se excluyen del análisis.

Nota: Los valores especificados deben estar en la variable, de lo contrario se muestra un mensaje de error para indicar que al menos uno de los grupos está vacío.

Análisis ROC: Definir grupos (numérico)

Para las variables de agrupación numérica, defina los dos grupos para la prueba *t* especificando dos valores, un punto medio o un punto de corte.

Nota: Los valores especificados deben estar en la variable, de lo contrario se muestra un mensaje de error para indicar que al menos uno de los grupos está vacío.

- **Usar valores especificados.** Especifique un valor para el grupo 1 y otro valor para el grupo 2. Los casos con cualquier otro valor se excluyen del análisis. Los números no tienen que ser enteros (por ejemplo, 6,25 y 12,5 son válidos).
- **Utilizar valor de punto medio.** Cuando está seleccionado, los grupos se separan en < y valores de punto medio.
- **Utilizar punto de corte.**
  - **Punto de corte.** Escriba un número que divida los valores de la variable de agrupación en dos conjuntos. Todos los casos con valores menores que el punto de corte forman un grupo y los casos con valores mayores o iguales que el punto de corte forman el otro grupo.

Curva ROC

Este procedimiento es un método útil para evaluar la realización de esquemas de clasificación en los que exista una variable con dos categorías por las que se clasifiquen los sujetos.

Nota: Versión29.0.1 en adelante, los menús y cuadros de diálogo del comando ROC se eliminan y este comando está disponible solo como sintaxis.

Ejemplo. Un banco tiene interés en clasificar a sus clientes dependiendo de si se retrasarán o no en el pago de sus préstamos; por tanto, se desarrollan métodos especiales para tomar estas decisiones. Las curvas COR se pueden utilizar para evaluar el grado de acierto de estos métodos.

Estadísticas. Es un área situada bajo la curva COR con un intervalo de confianza y puntos de coordenadas de la curva COR. Gráficos: Curva COR.

Métodos. Se puede calcular la estimación del área situado bajo la curva COR de forma paramétrica o no paramétrica mediante un modelo exponencial binegativo.

Curvas COR: Consideraciones sobre los datos

El antiguo procedimiento de Curva ROC soporta la inferencia estadística sobre una sola curva ROC. Esto también puede ser recuperado por el nuevo procedimiento de análisis ROC. Además, el nuevo procedimiento de análisis ROC puede comparar dos curvas ROC generadas a partir de grupos independientes o sujetos emparejados.

Datos. Las variables de contraste son cuantitativas. Las variables de contraste suelen estar constituidas por probabilidades, resultantes de un análisis discriminante o de una regresión logística, o bien compuestas por puntuaciones atribuidas en una escala arbitraria que indican el «grado de convicción» que tiene un evaluador de que el sujeto pueda pertenecer a una u otra categoría. La variable de estado puede ser de cualquier tipo e indicar la categoría real a la que pertenece un sujeto. El valor de la variable de estado indica la categoría que se debe considerar *positiva*.

Supuestos. Se considera que los números ascendentes de la escala del evaluador representan la creciente convicción de que el sujeto pertenece a una categoría. Por el contrario, los números descendentes representan la creciente convicción de que el sujeto pertenece a la otra categoría. El usuario deberá elegir qué dirección es *positiva*. También se considera que se conoce la categoría *real* a la que pertenece el sujeto.

Para obtener una curva ROC

1. En los menús elija: **Analizar** > **Clasificar** > **Análisis de la República de China...**

2. Seleccione uno o más **Variable de prueba** .
3. Seleccione uno **Variable de estado** .
4. Identifique el valor de *positivo* para la variable de estado.
5. Hacer clic **Mostrar** , seleccionar **Curva ROC** y **Con línea de referencia diagonal** , si es necesario.
6. Pulse **Continuar**

Curvas ROC: Opciones

Puede seleccionar las opciones siguientes para su análisis ROC:

Clasificación
Permite especificar si se debe incluir o excluir el valor del punto de corte al realizar una clasificación *positiva*. Este ajuste no afecta a los resultados.

Dirección de la prueba
Permite especificar la dirección de la escala según la categoría *positiva*.

Parámetros para el error estándar del área
Permite especificar el método de estimación del error estándar del área situada bajo la curva. Los métodos disponibles son exponenciales 'no paramétricos' y 'binegativos'. También se puede establecer el nivel para el intervalo de confianza. El rango disponible es entre el 50,1% y el 99,9%.

Valores perdidos
Permite especificar el tratamiento que reciben los valores perdidos.

Simulación

Los modelos predictivos, como una regresión lineal, requieren un conjunto de entradas conocidas para predecir un resultado o valor de destino. En muchas aplicaciones del mundo real, sin embargo, los valores de las entradas son inciertos. La simulación permite explicar la incertidumbre de las entradas en modelos predictivos y evaluar la posibilidad de varios resultados del modelo en presencia de esa incertidumbre. Por ejemplo, tiene un modelo de beneficio que incluye el coste de los materiales como una entrada, pero hay incertidumbre en ese coste por la volatilidad del mercado. Puede utilizar la simulación para modelar esa incertidumbre y determinar el efecto que tiene en los beneficios.

La simulación de IBM SPSS Estadística utiliza el método de Monte Carlo. Las entradas inciertas se modelan con distribuciones de probabilidad (como la distribución triangular), y los valores simulados de esas entradas se generan a partir de esas distribuciones. Las entradas cuyos valores se conocen se mantienen fijas en los valores conocidos. El modelo predictivo se evalúa utilizando un valor simulado para cada entrada incierta y los valores fijos de las entradas conocidas para calcular el destino (u destinos) del modelo. El proceso se repite muchas veces (normalmente decenas de miles o cientos de miles de veces), resultando en una distribución de los valores de destino que es posible utilizar para responder las preguntas de una naturaleza probabilística. En el contexto de IBM SPSS Estadística , cada repetición del proceso genera un caso diferente (registro) de datos que consiste en el conjunto de valores simulados de las entradas inciertas, los valores de las entradas fijas y el destino (o destinos) predichos del modelo.

También puede simular datos en ausencia de un modelo predictivo especificando distribuciones de probabilidad para las variables que se van a simular. Cada caso de datos generado consta del conjunto de valores simulados para las variables especificadas.

Para ejecutar una simulación, necesita especificar datos como el modelo predictivo, las distribuciones de probabilidad de las entradas inciertas, las correlaciones entre esas entradas y los valores de entradas fijas. Una vez haya especificado todos los detalles de una simulación, puede ejecutarla y, opcionalmente, guardar las especificaciones en un archivo de **plan de simulación**. Puede compartir el plan de simulación con otros usuarios, que pueden ejecutar la simulación sin necesidad de comprender los detalles de su creación.

Existen dos interfaces disponibles para trabajar con simulaciones. El Generador de simulaciones es una interfaz avanzada para usuarios que diseñan y ejecutan simulaciones. Proporciona el conjunto completo de funciones para diseñar una simulación, guardar las especificaciones en un archivo de plan de simulación, especificar los resultados y ejecutar la simulación. Puede crear una simulación

basada en un archivo de modelo de IBM SPSS, o en un conjunto de ecuaciones personalizadas que defina en el Generador de simulaciones. También puede cargar un plan de simulación existente en el Generador de simulaciones, modificar cualquiera de los ajustes y ejecutar la simulación, con la opción adicional de guardar el plan actualizado. Para los usuarios que tengan un plan de simulación y desean preferentemente ejecutar la simulación, existe una interfaz más simple. Permite modificar ajustes que permiten ejecutar la simulación en condiciones diferentes, pero no proporciona todas las funciones del Generador de simulaciones para el diseño de simulaciones.

Para diseñar una simulación basada en un archivo de modelo

1. En los menús seleccione:

   Analizar > **Simulación...**
2. Pulse en **Seleccionar Archivo modelo SPSS** y en **Continuar**.
3. Abra el archivo de modelo.

   Un modelo de archivo es un archivo XML que contiene el PMML de modelo creado desde IBM SPSS Estadística o IBM SPSS Modeler. Consulte el tema “Pestaña Modelo ” en la página 296 para obtener más información.
4. En la pestaña Simulación (del Generador de simulaciones), especifique las distribuciones de probabilidad de las entradas simuladas y los valores de entradas fijas. Si el conjunto de datos activo contiene datos históricos de entradas simuladas, pulse en **Ajustar todas** para determinar automáticamente la distribución que se ajusta mejor para cada entrada, y para determinar las correlaciones entre ellas. Para cada entrada simulada que no se ajuste a los datos históricos, debe especificar explícitamente una distribución seleccionando un tipo de distribución e introduciendo los parámetros necesarios.
5. Pulse en **Ejecutar** para ejecutar la simulación. De forma predeterminada, el plan de simulación, que especifica los detalles de la simulación, se guarda en la ubicación especificada en los ajustes de almacenamiento.

Se encuentran disponibles las siguientes opciones:

- Modificar la ubicación del plan de simulación guardado.
- Especificar correlaciones conocidas entre entradas simuladas.
- Calcular automáticamente una tabla de contingencia de asociaciones entre entradas categóricas y utilizar dichas asociaciones cuando los datos se generan para esas entradas.
- Especificar los análisis de sensibilidad para investigar el efecto de variar el valor de una entrada fija o variar un parámetro de distribución de una entrada simulada.
- Especificar opciones avanzadas como, por ejemplo, establecer el número máximo de casos para generar o solicitar muestreos de cola.
- Personalizar los resultados.
- Guardar los datos simulados en un archivo de datos.

Para diseñar una simulación basada en ecuaciones personalizadas

1. En los menús seleccione:

   Analizar > **Simulación...**
2. Pulse **Escribir las ecuaciones** y en **Continuar**.
3. Pulse en **Nueva ecuación** en la pestaña Modelo (del Generador de simulaciones) para definir cada ecuación del modelo predictivo.
4. Pulse la pestaña Simulación y especifique las distribuciones de probabilidad de entradas simuladas y valores de entradas fijas. Si el conjunto de datos activo contiene datos históricos de entradas simuladas, pulse en **Ajustar todas** para determinar automáticamente la distribución que se ajusta mejor para cada entrada, y para determinar las correlaciones entre ellas. Para cada entrada simulada

que no se ajuste a los datos históricos, debe especificar explícitamente una distribución seleccionando un tipo de distribución e introduciendo los parámetros necesarios.

5. Pulse en **Ejecutar** para ejecutar la simulación. De forma predeterminada, el plan de simulación, que especifica los detalles de la simulación, se guarda en la ubicación especificada en los ajustes de almacenamiento.

Se encuentran disponibles las siguientes opciones:

- Modificar la ubicación del plan de simulación guardado.
- Especificar correlaciones conocidas entre entradas simuladas.
- Calcular automáticamente una tabla de contingencia de asociaciones entre entradas categóricas y utilizar dichas asociaciones cuando los datos se generan para esas entradas.
- Especificar los análisis de sensibilidad para investigar el efecto de variar el valor de una entrada fija o variar un parámetro de distribución de una entrada simulada.
- Especificar opciones avanzadas como, por ejemplo, establecer el número máximo de casos para generar o solicitar muestreos de cola.
- Personalizar los resultados.
- Guardar los datos simulados en un archivo de datos.

Diseñar una simulación sin un modelo predictivo

1. En los menús, seleccione:

   Analizar > **Simulación...**

2. Pulse **Crear datos simulados** y pulse **Continuar**.
3. En la pestaña Modelo (en el Generador de simulaciones), seleccione los campos que desea simular. Puede seleccionar los campos del conjunto de datos activo o puede definir campos nuevos pulsando **Nuevo**.
4. Pulse la pestaña Simulación y especifique las distribuciones de probabilidad para los campos que se van a simular. Si el conjunto de datos activo contiene datos históricos para cualquiera de estos campos, pulse **Ajustar todo** para determinar la distribución que más se ajusta a todos los datos y determinar las correlaciones entre los campos. Para los campos que no se ajustan a los datos históricos, debe especificar de forma explícita una distribución seleccionando un tipo de distribución y especificando los parámetros necesarios.
5. Pulse en **Ejecutar** para ejecutar la simulación. De forma predeterminada, los datos simulados se guardan en el conjunto de datos nuevo especificado en la Configuración de guardar. Además, el plan de simulación, que especifica los detalles de la simulación, se guarda en la ubicación especificada en la Configuración de guardar.

Se encuentran disponibles las siguientes opciones:

- Modificar la ubicación para los datos simulados o el plan de simulación guardado.
- Especificar correlaciones conocidas entre campos simulados.
- Calcular automáticamente una tabla de contingencia de asociaciones entre campos categóricos y utilizar dichas asociaciones cuando se generan los datos para dichos campos.
- Especificar los análisis de sensibilidad para investigar el efecto de variar un parámetro de distribución para un campo simulado.
- Especificar opciones avanzadas como, por ejemplo, establecer el número de casos para generar.

Para ejecutar una simulación de un plan de simulación

Existen dos opciones disponibles para ejecutar una simulación a partir de un plan de simulación. Puede utilizar el cuadro de diálogo Ejecutar simulación, que está diseñado para ejecutarse a partir de un plan de simulación, o bien puede utilizar el Generador de simulaciones.

Para utilizar el cuadro de diálogo Ejecutar simulación:

1. En los menús seleccione:

   Analizar > **Simulación...**
2. Pulse en **Abrir un plan de simulación existente**.
3. Asegúrese de que la casilla de verificación **Abrir en el Generador de simulaciones** no está seleccionada y pulse **Continuar**.
4. Abra el plan de simulación.
5. Pulse en **Ejecutar** en el cuadro de diálogo Ejecutar simulación.

Para ejecutar la simulación del Generador de simulaciones:

1. En los menús seleccione:

   Analizar > **Simulación...**
2. Pulse en **Abrir un plan de simulación existente**.
3. Seleccione la casilla de verificación **Abrir en el Generador de simulaciones** y pulse en **Continuar**.
4. Abra el plan de simulación.
5. Modifique los ajustes que desee en la pestaña Simulación.
6. Pulse en **Ejecutar** para ejecutar la simulación.

También puede intentar lo siguiente:

- Configurar o modificar los análisis de sensibilidad para investigar el efecto de variar el valor de una entrada fija o variar un parámetro de distribución de una entrada simulada.
- Reajustar las distribuciones y correlaciones de entradas simuladas en nuevos datos.
- Cambiar la distribución de una entrada simulada.
- Personalizar los resultados.
- Guardar los datos simulados en un archivo de datos.

Generador de simulaciones

El Generador de simulaciones proporciona la gama completa de funciones para diseñar y ejecutar simulaciones. Permite ejecutar las siguientes tareas generales:

- Diseñar y ejecutar una simulación de un modelo de IBM SPSS definido en un archivo de modelo PMML.
- Diseñar y ejecutar una simulación de un modelo predictivo definido por un conjunto de ecuaciones personalizadas que especifique.
- Diseñar y ejecutar una simulación que genera datos en ausencia de un modelo predictivo.
- Ejecutar una simulación basada en un plan de simulación existente, pudiendo modificar los ajustes del plan.

Pestaña Modelo

Para simulaciones basadas en un modelo predictivo, la pestaña Modelo especifica el origen del modelo. Para simulaciones que no incluyen un modelo predictivo, la pestaña Modelo especifica los campos que se van a simular.

Seleccione un archivo de modelo SPSS. Esta opción especifica que el modelo predictivo se define en un archivo de modelo IBM SPSS. Un archivo de modelo de IBM SPSS es un archivo XML o un archivador de archivos comprimidos (archivo .zip) que contiene el PMML de modelo creado desde IBM SPSS Estadística o IBM SPSS Modeler. Los modelos predictivos están creados por procedimientos, como Regresión lineal y Árboles de decisión en IBM SPSS Estadística , y se puede exportar a un archivo de modelo. Puede utilizar un archivo de modelo diferente pulsando en **Examinar** y desplazándose al archivo que desee.

Modelos PMML admitidos por Simulación

- Regresión lineal

- Modelo lineal automático
- Modelo lineal generalizado
- Modelo mixtos lineales generalizados
- Modelo lineal general
- Regresión logística binaria
- Regresión logística multinomial
- Regresión multinomial ordinal
- Regresión de Cox
- Árbol
- Árbol potenciado (C5)
- Discriminante
- Clúster de dos fases
- Clúster de K-medias
- Red neuronal
- Conjunto de reglas (Lista de decisiones)

Nota:

- No se admite en Simulación el uso de modelos PMML que tengan múltiples campos de destino (variables) o divisiones.
- Los valores de entradas de cadena en modelos de regresión logística binaria están limitados a bytes en el modelo. Si está adaptando tales entradas de cadena al conjunto de datos activo, asegúrese de que los valores de los datos no superan 8 bytes de longitud. Los valores de datos que superan 8 bytes se excluyen de la distribución categórica asociada de la entrada, y aparecen sin coincidencias en la tabla de salida Categorías sin coincidencias.

Escriba las ecuaciones para el modelo. Esta opción especifica que el modelo predictivo se compone de una o más ecuaciones personalizadas que puede crear el usuario. Cree las ecuaciones pulsando en **Nueva ecuación**. Se abrirá el Editor de ecuaciones. Puede modificar ecuaciones existentes, copiarlas para utilizarlas como plantillas de nuevas ecuaciones, reordenarlas y eliminarlas.

- El Generador de simulaciones no admite sistemas de ecuaciones simultáneas o ecuaciones que no son lineales en la variable destino.
- Las ecuaciones personalizadas se evalúan en el orden en que se especifican. Si la ecuación de un destino especificado depende de otro destino, el otro destino se debe definir mediante una ecuación anterior.

  Por ejemplo, teniendo en cuenta el conjunto de ecuaciones siguiente, la ecuación de *beneficios* depende de los valores de *ingresos* y *gastos*, por lo que las ecuaciones de *ingresos* y *gastos* deben anteceder a la ecuación de *beneficios*.

```
revenue = price*volume

expenses = fixed + volume*(unit_cost_materials + unit_cost_labor)

profit = revenue - expenses
```

Crear datos simulados sin un modelo. Seleccione esta opción para simular datos sin un modelo predictivo. Especifique los campos que deben simularse seleccionando los campos del conjunto de datos activo o pulsando **Nuevo** para definir campos nuevos.

Editor de ecuaciones

El Editor de ecuaciones permite crear o modificar una ecuación personalizada de su modelo predictivo.

- La expresión de la ecuación puede contener campos del conjunto de datos activo o nuevos campos de entrada que define en el Editor de ecuaciones.

- Puede especificar propiedades del destino como su nivel de medición, etiquetas de valores y si los resultados están generados para el destino.
- Puede utilizar los destinos de ecuaciones definidas anteriormente como entradas de la ecuación actual, permitiéndole crear ecuaciones acopladas.
- Puede adjuntar un comentario descriptivo a la ecuación. Los comentarios se muestran junto con la ecuación de la pestaña Modelo.

1. Introduzca el nombre del destino. Opcionalmente, pulse en **Editar** en el cuadro de texto Destino para abrir el cuadro de diálogo Entradas definidas, permitiendo cambiar las propiedades predeterminadas del destino.
2. Para crear una expresión, puede pegar los componentes en el campo Expresión numérica o escribir directamente en dicho campo.

- Puede crear su expresión utilizando campos del conjunto de datos activo o puede definir nuevas entradas pulsando en el botón **Nuevo**. Se abrirá el cuadro de diálogo Definir entradas.
- Puede pegar las funciones seleccionando un grupo de la lista Grupo de funciones y pulsando dos veces en la función de la lista de funciones (o seleccione la función y pulse en la flecha que se encuentra sobre la lista Grupo de funciones). Especifique los parámetros indicados por los signos de interrogación. El grupo de funciones con la etiqueta **Todo** contiene una lista de todas las funciones variables. En un área reservada del cuadro de diálogo se muestra una breve descripción de la función actualmente seleccionada.
- Las constantes de cadena deben ir entre comillas.
- Si los valores contienen decimales, debe utilizarse una coma(,) como indicador decimal.

Nota: la simulación no admite ecuaciones personalizadas con destinos de cadena.

Entradas definidas

El cuadro de diálogo Entradas definidas permite definir nuevas entradas y las propiedades de los destinos.

- Si una entrada que se va a utilizar en una ecuación no existe en el conjunto de datos activo, debe definirlo para que se pueda utilizar en la ecuación.
- Si simula datos con un modelo predictivo, debe definir todas las entradas simuladas que no existan en el conjunto de datos activo.

Nombre. Especifique el nombre del destino o la entrada.

Destino. Puede especificar el nivel de medición de un destino. El nivel de medición predeterminado es continuo. También puede especificar si se creará el resultado para este destino. Por ejemplo, en el caso de un grupo de ecuaciones acopladas es posible que solo le interese el resultado del destino de la ecuación final, por lo que suprimiría el resultado del resto de destinos.

Entrada que se simulará. Especifica que los valores de la entrada se simularán de acuerdo con una distribución de probabilidad especificada (la distribución de probabilidad se especifica en la pestaña Simulación). El nivel de medición determina el conjunto predeterminado de distribuciones que se consideran al encontrar la mejor distribución de la entrada (pulsando **Ajustar** o **Ajustar todas** en la pestaña de Simulación). Por ejemplo, si el nivel de medición es continuo, la distribución normal (adecuada para datos continuos) se consideraría, pero la distribución binomial no.

Nota: Seleccione el nivel de medición de Cadena para las entradas de cadena. Las entradas de cadena que se simularán están restringidas a la distribución categórica.

Entrada de valor fijo. Especifica que se conoce el valor de la entrada y que se fijará en el valor conocido. Las entradas fijas pueden ser numéricas o de cadena. Especifique un valor para la entrada fija. Los valores de cadena no deben ir entre comillas.

Etiquetas de valor. Puede especificar etiquetas de valores de destino, entradas simuladas y fijas. Las etiquetas de valores se utilizan en gráficos y tablas de resultados.

Pestaña Simulación

La pestaña Simulación especifica todas las propiedades de la simulación diferentes a la del modelo predictivo. Puede ejecutar las siguientes tareas generales en la pestaña Simulación:

- Especificar las distribuciones de probabilidad de las entradas simuladas y valores de entradas fijas.
- Especificar correlaciones entre entradas simuladas. Para entradas categóricas, puede especificar que se utilicen las asociaciones que existen entre las entradas del conjunto de datos activos cuando se generen datos para dichas entradas.
- Especificar opciones avanzadas como muestreos de cola y criterios para ajustar distribuciones a datos históricos.
- Personalizar los resultados.
- Especificar la ubicación en la que guardar el plan de simulación y opcionalmente, guardar los datos simulados.

Campos simulados

Para ejecutar una simulación, cada campo de entrada debe especificarse como fijo o simulado. Las entradas simuladas son aquellas cuyos valores son inciertos y se generarán a partir de una distribución de probabilidad especificada. Cuando los datos históricos están disponibles para que se simulen las entradas, las distribuciones que se ajusten mejor a los datos se podrán determinar de forma automática, junto a las correlaciones entre dichas entradas. También puede especificar manualmente las distribuciones o correlaciones si los datos históricos no están disponibles o si necesita distribuciones o correlaciones específicas.

Las entradas fijas son aquellas cuyos valores se conocen y permanecen constantes en cada caso generado en la simulación. Por ejemplo, tiene un modelo de regresión lineal de ventas como una función de un número de entradas incluyendo el precio y desea mantener el precio fijo en el mercado actual. A continuación especificaría el precio como entrada fija.

Para simulaciones basadas en un modelo predictivo, cada predictor del modelo es un campo de entrada para la simulación. Para simulaciones que no incluyen un modelo predictivo, los campos que se especifican en la pestaña Modelo son los campos para la simulación.

Ajuste automático de distribuciones y cálculo de correlaciones de entradas simuladas. Si la base de datos activa contiene datos históricos de las entradas que desee simular, puede encontrar automáticamente las distribuciones que mejor se ajustan a esas entradas, así como determinar las correlaciones entre ellas. Los pasos son los siguientes:

1. Compruebe que todas las entradas que desea simular coinciden con el campo correcto en el conjunto de datos activo. Las entradas se listan en la columna Entrada y la columna Ajustar muestra el campo correspondiente en el conjunto de datos activos. Puede relacionar una entrada con un campo diferente en el conjunto de datos activo seleccionando un elemento diferente en la lista desplegable Ajustar a.

   Un valor de *-Ninguno-* en la columna Ajustar a indica que la entrada no se ha podido relacionar automáticamente a un campo en el conjunto de datos activo. De forma predeterminada, las entradas se relacionan con los campos del conjunto en el nombre y el nivel de medición y tipo (numéricas o de cadena). Si el conjunto de datos activo no contiene datos históricos de la entrada, especifique manualmente la distribución de la entrada o especifique la entrada como entrada fija, tal y como se describe a continuación.
2. Pulse **Ajustar todas**.

La distribución con mejor ajuste y sus parámetros asociados se muestran en la columna Distribución junto con una representación de la distribución superpuesta en un histograma (o gráfico de barras) de los datos históricos. Las correlaciones entre las entradas simuladas se muestran en los ajustes de correlaciones. Puede examinar los resultados de ajuste y personalizar el ajuste de distribución automático de una entrada concreta seleccionando la fila de la entrada y pulsando en **Detalles del ajuste**. Consulte el tema “Detalles del ajuste ” en la página 301 para obtener más información.

Puede ejecutar el ajuste de distribución automático de una entrada concreta seleccionando la fila de la entrada y pulsando en **Ajustar**. Las correlaciones de todas las entradas simuladas que corresponden con los campos del conjunto de datos activo también se calculan automáticamente.

Nota:

- Los casos con valores que faltan para cualquier entrada simulada se excluyen del ajuste de distribución, el cálculo de correlaciones y el cálculo de la tabla de contingencia opcional (para entradas con una distribución categórica). De forma opcional, puede especificar si los valores que faltan del usuario de entradas con una distribución categórica se tratan como válidos. De forma predeterminada, se tratan como valores que faltan. Para obtener más información, consulte el tema “Opciones avanzadas ” en la página 303.
- Para las entradas continuas y ordinales, si no se puede encontrar ningún ajuste aceptable para cualquiera de las distribuciones probadas, se sugiere la distribución empírica como el ajuste más cercano. En entradas continuas, la distribución empírica es la función de distribución acumulada de los datos históricos. En entradas ordinales, la distribución empírica es la distribución categórica de los datos históricos.

Distribuciones de especificación manuales. Puede especificar manualmente la distribución de probabilidad de cualquier entrada simulada seleccionando la distribución de la lista desplegable **Tipo** e introduciendo los parámetros de distribución de la cuadrícula Parámetros. Una vez haya introducido los parámetros de una distribución, un gráfico de muestra de la distribución, basado en los parámetros específicos, se mostrará junto a la cuadrícula Parámetros. A continuación se incluyen algunas notas sobre distribuciones concretas:

- **Categóricas.** La distribución categórica describe un campo de entrada que tiene un número fijo de valores, referido como categorías. Cada categoría tiene una probabilidad asociada, de forma que la suma de las probabilidades de todas las categorías es igual a uno. Para especificar una categoría, pulse con el botón derecho del ratón en la columna de la izquierda de la cuadrícula Parámetros y especifique el valor de la categoría. Especifique la probabilidad asociada con la categoría en la columna derecha.

  Nota: Las entradas categóricas de un modelo PMML tienen categorías que se determinan a partir del modelo y no se pueden modificar.
- **Binomial negativa - Errores.** Describe la distribución del número de fallos en una secuencia de ensayos antes de detectar un número especificado de éxitos. El parámetro *thresh* es el número especificado de éxitos y el parámetro *prob* es la probabilidad de éxito en cualquier prueba.
- **Binomial negativa - Intentos.** Describe la distribución del número de ensayos necesarios para detectar un número especificado de éxitos. El parámetro *thresh* es el número especificado de éxitos y el parámetro *prob* es la probabilidad de éxito en cualquier prueba.
- **Rango.** Esta distribución consta de un conjunto de intervalos con una probabilidad asignada a cada intervalo de tal forma que la suma de las probabilidades en todos los intervalos es igual a 1. Los valores de un intervalo determinado se extraen de una distribución uniforme definida en ese intervalo. Los intervalos se especifican introduciendo un valor mínimo, un valor máximo y una probabilidad asociada.

  Por ejemplo, usted cree que el coste de una materia prima tiene un 40% de posibilidades de caer entre 10 - 15 dólares por unidad y un 60% de posibilidades de caer entre 15 - 20 dólares por unidad. Modelaría el coste con una distribución del rango de los dos intervalos [10 - 15] y [15 - 20], definiendo la probabilidad asociada con el primer intervalo a 0,4 y la probabilidad asociada con el segundo intervalo a 0,6. Los intervalos no tienen que ser contiguos y se pueden superponer. Por ejemplo, podría especificar los intervalos 10 - 15 y 20 - 25 o 10 - 15 y 13 - 16.
- **Weibull.** El parámetro *c* es un parámetro de ubicación opcional, que especifica dónde se encuentra el origen de la distribución.

Los parámetros de las siguientes distribuciones tienen el mismo significado en las funciones variables aleatorias asociadas disponibles en el cuadro de diálogo Calcular variable: Bernoulli, Beta, Binomial, Exponencial, Gamma, Lognormal, Binomial negativa (Intentos y Fallos), Normal, Poisson y Uniforme.

Especificación de entradas fijas. Especifica una entrada fija seleccionando Fija en la lista desplegable **Tipo** de la columna Distribución, e introduciendo el valor fijo. El valor puede ser numérico o cadena, dependiendo de si la entrada es numérica o de cadena. Los valores de cadena no deben ir entre comillas.

Especificación de límites en valores simulados. La mayoría de distribuciones admiten la especificación de los límites superiores e inferiores de los valores simulados. Puede especificar un límite inferior introduciendo un valor en el cuadro de texto **Mín** y especificar un límite superior introduciendo un valor en el cuadro de texto **Máx**.

Bloqueo de entradas. El bloqueo de una entrad, seleccionando la casilla de verificación en la columna con el icono de bloqueo, excluye la entrada del ajuste de distribución automático. Esto es muy útil cuando especifica manualmente una distribución o un valor fijo y desea garantizar que no se verá afectada por el ajuste de distribución automático. El bloqueo también es útil si desea compartir el plan de simulación con los usuarios que lo ejecutarán en el diálogo Ejecutar simulación y desea evitar los cambios en algunas entradas. En este sentido, las especificaciones para entradas bloqueadas no se pueden modificar en el diálogo Ejecutar simulación.

Análisis de sensibilidad. El análisis de sensibilidad le permite investigar el efecto de los cambios sistemáticos en una entrada fija o en un parámetro de distribución para una entrada simulada generando un conjunto independiente de casos simulados -efectivamente, una simulación separada- para cada valor especificado. Para especificar los análisis de sensibilidad, seleccione una entrada fija o simulada y pulse en **Análisis de sensibilidad**. El análisis de sensibilidad está limitado a una única entrada fija o un parámetro de distribución único de una entrada simulada. Consulte el tema “Análisis de sensibilidad ” en la página 302 para obtener más información.

Iconos de estado de Ajustar

Los iconos de la columna Ajustar a indican el estado de ajuste de cada campo de entrada.

Tabla 8. Iconos de estado

Icono	Descripción
	No se ha especificado una distribución para la entrada y la entrada no se ha especificado como fija. Para ejecutar la simulación, debe especificar una distribución para esta entrada o definirla como fija y especificar el valor fijo.
	La entrada se ha fijado anteriormente a un campo que no existe en el conjunto de datos activo. No se necesita ninguna acción salvo que desee volver a ajustar la distribución de la entrada al conjunto de datos activo.
	La mejor distribución de ajuste se ha sustituido por una distribución diferente del cuadro de diálogo Detalles del ajuste.
	La entrada se define a la mejor distribución de ajuste.
	La distribución se ha especificado manualmente o se han especificado las iteraciones del análisis de sensibilidad para esta entrada.

Detalles del ajuste

El cuadro de diálogo Detalles del ajuste muestra los resultados del ajuste de distribución automático de una entrada concreta. Las distribuciones se ordenan por la bondad del ajuste, apareciendo en primer lugar la distribución con mejor ajuste. Puede sustituir la distribución de mejor ajuste seleccionando el botón de opción de la distribución que desee en la columna Utilizar. Al seleccionar un botón de opción en la columna Utilizar también muestra un gráfico de distribución superpuesto en un histograma (o gráfico de barras) de los datos históricos de esa entrada.

Estadísticos de ajuste. De forma predeterminada, en los campos continuos, la prueba Anderson-Darling se utiliza para determinar su idoneidad. Del mismo modo, y para los campos continuos únicamente, puede especificar la prueba Kolmogorov-Smirnoff de idoneidad, seleccionando esa opción en los ajustes Opciones avanzadas. En entradas continuas, los resultados de ambas pruebas se muestran en la columna Estadísticos de ajuste (A para Anderson-Darling y K para Kolmogorov-Smirnoff), usando la prueba elegida

para ordenar las distribuciones. En las entradas ordinales y nominales se utiliza la comprobación de chi-cuadrado. También se muestran los valores p asociados a las pruebas.

Parámetros. Los parámetros de distribución asociados con cada distribución ajustada se muestran en la columna Parámetros. Los parámetros de las siguientes distribuciones tienen el mismo significado en las funciones variables aleatorias asociadas disponibles en el cuadro de diálogo Calcular variable: Bernoulli, Beta, Binomial, Exponencial, Gamma, Lognormal, Binomial negativa (Intentos y Fallos), Normal, Poisson y Uniforme. En la distribución categórica, los nombres de los parámetros son las categorías y los valores de parámetros son las probabilidades asociadas.

Reajuste con un conjunto de distribución personalizado. De forma predeterminada, el nivel de medición de la entrada se utiliza para determinar el conjunto de distribuciones consideradas para el reajuste de distribución automático. Por ejemplo, las distribuciones continuas como lognormal y gamma se consideran cuando se ajusta una entrada continua; pero las distribuciones discretas como Poisson y binomial, no. Puede elegir un subconjunto de las distribuciones predeterminado seleccionando las distribuciones en la columna Reajustar. También puede sustituir el conjunto predeterminado de distribuciones seleccionando un nivel de medición diferente en la lista desplegable **Tratar como (Medida)** y seleccionando las distribuciones en la columna Reajustar. Pulse en **Ejecutar reajuste** para volver a ajustar con el conjunto de distribución personalizado.

Nota:

- Los casos con valores que faltan para cualquier entrada simulada se excluyen del ajuste de distribución, el cálculo de correlaciones y el cálculo de la tabla de contingencia opcional (para entradas con una distribución categórica). De forma opcional, puede especificar si los valores que faltan del usuario de entradas con una distribución categórica se tratan como válidos. De forma predeterminada, se tratan como valores que faltan. Para obtener más información, consulte el tema “Opciones avanzadas ” en la página 303.
- Para las entradas continuas y ordinales, si no se puede encontrar ningún ajuste aceptable para cualquiera de las distribuciones probadas, se sugiere la distribución empírica como el ajuste más cercano. En entradas continuas, la distribución empírica es la función de distribución acumulada de los datos históricos. En entradas ordinales, la distribución empírica es la distribución categórica de los datos históricos.

Análisis de sensibilidad

Los análisis de sensibilidad permiten investigar el efecto de modificar una entrada fija o un parámetro de distribución de una entrada simulada en un conjunto específico de valores. Se genera un conjunto independiente de casos simulados (una simulación separada) para cada valor especificado, lo que le permite investigar el efecto de modificar la entrada. Cada conjunto de casos simulados se denomina **iteración**.

Iterar. Esta opción permite especificar el conjunto de valores en el que se modificará la entrada.

- Si modifica el valor de un parámetro de distribución, seleccione el parámetro de la lista desplegable. Introduzca el conjunto de valores en el valor Parámetro mediante la cuadrícula de iteraciones. Al pulsar en **Continuar** se añadirán los valores especificados en la cuadrícula Parámetros de la entrada asociada, con un índice que especifica el número de iteración del valor.
- En las distribuciones categóricas y de rango, las probabilidades de las categorías o intervalos respectivamente se pueden variar, pero los valores de las categorías y los puntos finales de los intervalos no se pueden modificar. Seleccione una categoría o intervalo en la lista desplegable y especifique el conjunto de probabilidades en el valor del parámetro mediante la cuadrícula de iteración. Las probabilidades de otras categorías o intervalos se ajustarán automáticamente.

Sin iteraciones. Utilice esta opción para cancelar las iteraciones de una entrada. Si pulsa en **Continuar** eliminará las iteraciones.

Correlaciones

Los campos de entrada que se van a simular con frecuencia se sabe que están correlacionados como, por ejemplo, altura y ponderación. Las correlaciones entre las entradas que se simularán se deben tener en cuenta para garantizar que los valores simulados mantengan esas correlaciones.

Recalcular correlaciones al ajustar. Esta opción especifica que las correlaciones entre las entradas simuladas se calculen automáticamente al ajustar las distribuciones en el conjunto de datos activo mediante las acciones **Ajustar todas** o **Ajustar** en los ajustes de Campos simulados.

No recalcular correlaciones al ajustar. Seleccione esta opción si desea especificar manualmente las correlaciones y evitar que se sobrescriban al ajustar automáticamente distribuciones en el conjunto de datos activo. Los valores que se especifican en la cuadrícula Correlaciones deben estar entre -1 y 1. Un valor de 0 especifica que no hay correlación entre el par asociado de entradas.

Restablecer. Restablece todas las correlaciones a 0.

Utilizar tabla de contingencia de varios factores ajustada para entradas con una distribución categórica. Para entradas con una distribución categórica, puede calcular automáticamente una tabla de contingencia de varios factores a partir del conjunto de datos activo que describe las asociaciones entre dichas entradas. La tabla de contingencia se utiliza cuando se generan los datos para dichas entradas. Si opta por guardar el plan de simulación, la tabla de contingencia se guarda en el archivo de plan y se utilizan cuando se ejecuta el plan.

- **Calcular tabla de contingencia a partir del conjunto de datos activo.** Si trabaja con un plan de simulación existente que contiene una tabla de contingencia, puede volver a calcular la tabla de contingencia a partir del conjunto de datos activo. Esta acción sustituye la tabla de contingencia del archivo de plan cargado.
- **Utilizar tabla de contingencia del plan de simulación cargado.** De forma predeterminada, cuando carga un plan de simulación que contiene una tabla de contingencia, se utiliza la tabla del plan. Puede volver a calcular la tabla de contingencia a partir del conjunto de datos activo seleccionando **Calcular tabla de contingencia a partir del conjunto de datos activo**.

Opciones avanzadas

Número máximo de casos. Especifica el número máximo de casos de datos simulados y los valores de destinos asociados que se generarán. Si se especifica el análisis de sensibilidad, es el número máximo de casos de cada iteración.

Destino de criterios de parada. Si su modelo predictivo contiene más de un destino, podrá seleccionar el destino al que se aplican los criterios de parada.

Criterios de parada. Estas opciones especifican los criterios para detener la simulación, potencialmente antes de generar el número máximo de casos permitidos.

- **Continuar hasta alcanzar el máximo.** Especifica que los casos simulados se generarán hasta que se alcance el número máximo de casos.
- **Detener cuando las colas se hayan muestreado.** Utilice esta opción si desea asegurarse de que una de las colas de una distribución de destino especificada se ha muestreado correctamente. Los casos simulados se generarán hasta que se complete el muestreo de la cola específica o se alcance el número máximo de casos. Si su modelo predictivo contiene múltiples destinos, seleccione el destino al que se aplicarán los criterios en la lista desplegable **Destino de criterios de parada**.

  Tipo. Puede definir el límite de la región de cola especificando un valor de destino como 10.000.000 o un percentil como el 99º. Si selecciona Valor en la lista desplegable **Tipo**, introduzca el valor del límite en el cuadro de texto Valor y utilice la lista desplegable **Lado** para especificar si es el límite de la región de cola izquierda o la región de cola derecha. Si selecciona Percentil en la lista desplegable **Tipo**, introduzca un valor en el cuadro de texto Percentil.

  Frecuencia. Especifique el número de valores del destino que debe estar en la región para garantizar que la cola se ha muestreado correctamente. La generación de los casos se detendrá cuando se alcance este número.

- **Detener si el intervalo de confianza de la media está en el umbral especificado.** Utilice esta opción si desea asegurarse de que la media de un destino concreto se conoce con un grado especificado de precisión. Los casos simulados se generarán hasta que se alcance el grado específico de precisión o el número máximo de casos. Para utilizar esta opción, especifique un nivel de confianza y un umbral. Los casos simulados se generarán hasta que el intervalo de confianza asociado con el nivel de confianza esté dentro del umbral. Por ejemplo, puede utilizar esta opción para especificar que los casos se generan hasta que el intervalo de confianza de la media al 95% del nivel de confianza esté en el 5% del valor medio. Si su modelo predictivo contiene múltiples destinos, seleccione el destino al que se aplicarán los criterios en la lista desplegable **Destino de criterios de parada**.

  Tipo de umbral. Puede especificar el umbral como un valor numérico o como porcentaje de la media. Si selecciona Valor en la lista desplegable **Tipo de umbral**, introduzca el umbral en el cuadro de texto Umbral como valor. Si selecciona Porcentaje en la lista desplegable **Tipo de umbral**, introduzca un valor en el cuadro de texto Umbral como porcentaje.

Número de casos de la muestra. Especifica el número de casos que se utilizarán al ajustar automáticamente las distribuciones de las entradas simuladas en el conjunto de datos activo. Si su conjunto de datos es muy grande, puede que desee limitar el número de casos utilizados para el ajuste de distribución. Si selecciona **Limita a N casos**, se utilizarán los primeros N casos.

Bondad de criterios de ajuste (Continua). En entradas continuas, puede utilizar la prueba Anderson-Darling o la prueba Kolmogorov-Smirnoff de bondad de ajuste para ordenar distribuciones al ajustar distribuciones de entradas simuladas en el conjunto de datos activo. La prueba Anderson-Darling está seleccionada de forma predeterminada y está especialmente recomendada si desea garantizar el mejor ajuste posible en las regiones de cola.

Distribución empírica. En entradas continuas, la distribución empírica es la función de distribución acumulada de los datos históricos. Puede especificar el número de agrupaciones utilizadas para calcular las distribuciones empíricas de las entradas continuas. El valor predeterminado es 100 y el máximo es 1.000.

Replicar resultados. Al establecer una semilla aleatoria podrá replicar las simulaciones. Especifique un entero o pulse en **Generar**, lo que creará un entero pseudo-aleatorio entre 1 y 2147483647, ambos inclusive. El valor predeterminado es 629111597.

Nota: Para una semilla aleatoria determinada, los resultados se duplican a menos que cambie el número de hebras. En un sistema determinado, el número de hebras es constante a menos que lo cambie ejecutando la sintaxis de comando SET THREADS. El número de hebras puede cambiar si se ejecuta la simulación en un sistema diferente porque se utiliza un algoritmo interno para determinar el número de hebras para cada sistema.

Valores perdidos del usuario para entradas con una distribución categórica. Estos controles especifican si los valores perdidos del usuario de entradas con una distribución categórica se tratan como válidos. Los valores perdidos del sistema para todos los tipos de entrada se tratan siempre como no válidos. Todas las entradas deben tener valores válidos para que un caso se incluya en el ajuste de distribución, el cálculo de correlaciones y el cálculo de la tabla de contingencia opcional.

Funciones de densidad

Estos ajustes permiten personalizar los resultados de las funciones de densidad de probabilidad y las funciones de distribución acumuladas de destinos continuos, así como gráficos de barras de los valores pronosticados de destinos categóricos.

Función de densidad de probabilidad (PDF). La función de densidad de probabilidad muestra la distribución de valores de destino. En destinos continuos, permite determinar la probabilidad de que el destino esté dentro de una región concreta. En destinos categóricos (destinos con un nivel de medición del nominal u ordinal), se genera un gráfico de barras que muestra en porcentaje de casos que entran dentro de cada categoría del destino. Existen otras opciones para los destinos categóricos de modelos PMML disponibles con los valores de Categoría para los ajustes de informes que se describen a continuación.

Para modelos de clúster de dos fases y modelos de clúster de K-medias, se genera un gráfico de barras de la pertenencia a clústeres.

Función de distribución acumulada (CDF). La función de distribución acumulada muestra la probabilidad de que el valor del destino sea menor o igual que un valor especificado. Solo está disponible para destinos continuos.

Posiciones del deslizador. Puede especificar las posiciones iniciales de las líneas de referencia móviles en gráficos PDF y CDF. Los valores que se especifican para las líneas inferior y superior hacen referencia a las posiciones a lo largo del eje horizontal, no a percentiles. Puede eliminar la línea inferior seleccionando **-Infinity** o puede eliminar la línea superior seleccionando **Infinity**. De forma predeterminada, las líneas se sitúan en los percentiles 5 y 95. Cuando se muestran varias funciones de distribución en un único gráfico (debido a varios destinos o resultados de iteraciones del análisis de sensibilidad), el valor predeterminado hace referencia a la distribución de la primera iteración o del primer destino.

Líneas de referencia (Continuas). Puede solicitar varias líneas de referencia verticales para añadirlas a funciones de densidad de probabilidad y funciones de distribución acumulada para destinos continuos.

- **Sigmas.** Puede añadir líneas de referencia por encima o por debajo de un número especificado de desviaciones estándar desde la media de un objetivo.
- **Percentiles.** Ahora puede añadir líneas de referencia a uno o dos valores de los percentiles de la distribución de un objetivo introduciendo los valores en los cuadros de texto inferior y superior. Por ejemplo, un valor de 95 en el cuadro de texto superior representa el percentil 95, que es el valor por debajo del cual cae el 95% de las observaciones. Del mismo modo, un valor de 5 en el cuadro de texto inferior representa el percentil 5, que es el valor por debajo del cual cae el 5% de las observaciones.
- **Líneas de referencia personalizadas.** Puede añadir líneas de referencia a los valores especificados del destino.

Nota: Cuando se muestran varias funciones de distribución en un único gráfico (debido a varios destinos o resultados de iteraciones del análisis de sensibilidad), las líneas de referencia sólo se aplican a la distribución de la primera iteración o del primer destino. Puede añadir líneas de referencia a otras distribuciones desde el diálogo Opciones de gráfico, al que se accede desde el gráfico PDF o CDF.

Superponer resultados de destinos continuos diferentes. En el caso de múltiples destinos continuos, especifica si las funciones de distribución de todos estos destinos se muestran en un único gráfico, con un gráfico de funciones de densidad de probabilidad y otra para funciones de distribución acumuladas. Si esta opción no se selecciona, los resultados de cada destino se mostrarán en un gráfico diferente.

Valores de categoría que se incluirán en el informe. En modelos PMML con destinos categóricos, el resultado del modelo es un conjunto de probabilidades pronosticadas, una de cada categoría, en la que el valor de destino entra dentro de cada categoría. La categoría con la mayor probabilidad se toma como la categoría pronosticada y se utiliza en la generación del gráfico de barras que se describe en el ajuste **Función de densidad de probabilidad** anterior. Si selecciona **Categoría pronosticada** se generará el gráfico de barras. Si selecciona **Probabilidades pronosticadas** se generarán histogramas de la distribución de las probabilidades pronosticadas de cada una de las categorías del destino.

Agrupación de análisis de sensibilidad. Las simulaciones que incluyen análisis de sensibilidad generan un conjunto independiente de valores de destino pronosticados para cada iteración que define el análisis (una iteración para cada valor de la entrada que se está variando). Si existen iteraciones, el gráfico de barras de la categoría pronosticada de un destino categórico se muestra como un gráfico de barras clúster que incluye los resultados de todas las iteraciones. Puede seleccionar agrupar las categorías o agrupar las iteraciones.

Resultado

Gráficos de tornado. Los gráficos de tornado son gráficos de barra que muestran relaciones entre destinos y entradas simuladas utilizando una variedad de métricas.

- **Correlación del destino con la entrada.** Esta opción crea un gráfico de tornado de los coeficientes de correlación entre un destino especificado y cada una de sus entradas simuladas. Este tipo de gráfico de

tornado no admite destinos con un nivel de medición nominal u ordinal ni entradas simuladas con una distribución categórica.

- **Contribución a la varianza.** Esta opción crea un gráfico de tornado que muestra la contribución a la varianza de un destino de cada una de sus entradas simuladas, lo que permite evaluar el grado en el que cada entrada contribuye a la incertidumbre general del destino. Este tipo de gráfico de tornado no admite destinos con niveles de medición nominales u ordinales, o entradas simuladas con cualquiera de las distribuciones siguientes: categórica, Bernoulli, binomial, Poisson o binomial negativa.
- **Sensibilidad del destino para cambiar.** Esta opción crea un gráfico de tornado que muestra el efecto del destino del destino de modulación de cada entrada simulada más o menos un número especificado de desviaciones estándar de la distribución asociada con la entrada. Este tipo de gráfico de tornado no admite destinos con niveles de medición nominales u ordinales, o entradas simuladas con cualquiera de las distribuciones siguientes: categórica, Bernoulli, binomial, Poisson o binomial negativa.

Diagramas de cajas de distribuciones de destino. Los diagramas de cajas están disponibles para destinos continuos. Seleccione **Superponer resultados de destinos diferentes** si su modelo predictivo tiene múltiples destinos continuos y desea visualizar los diagramas de cajas de todos los destinos en un gráfico único.

Diagramas de dispersión de destinos frente a entradas. Los diagramas de dispersión frente a entradas simuladas están disponibles para destinos continuos y categóricos e incluyen dispersiones del destino con entradas continuas y categóricas. Las dispersiones que incluyen un destino o una entrada categórica se muestran como un mapa de calor.

Crear una tabla de valores percentiles. En destinos continuos, puede obtener una tabla de percentiles especificados de las distribuciones de destino. Los cuartiles (los percentiles 25, 50 y 75) dividen las observaciones en cuatro grupos de igual tamaño. Si desea un número igual de grupos que no sea cuatro, seleccione **Intervalos** y especifique el número. Seleccione **Percentiles personalizados** para especificar percentiles individuales, por ejemplo, el percentil 99.

Estadísticos descriptivos de distribuciones de destino. Esta opción crea tablas de estadísticos descriptivos de destinos continuos y categóricos así como entradas continuas. En destinos continuos, la tabla incluye la media, la desviación estándar, la mediana, el mínimo y el máximo, el intervalo de confianza de la media en el nivel especificado y los percentiles 5 y 95 de la distribución de destino. En destinos categóricos, la tabla incluye el porcentaje de casos que entran en cada categoría del destino. En destinos categóricos de modelos PMML, la tabla también incluye la probabilidad media de cada categoría del destino. En entradas continuas, la tabla incluye la media, desviación estándar, mínima y máxima.

Correlaciones y tabla de contingencia para entradas. Esta opción muestra una tabla de coeficientes de correlación entre entradas simuladas. Cuando se generan entradas con distribuciones categóricas a partir de una tabla de contingencia, también se muestra la tabla de contingencia de los datos que se generan para dichas entradas.

Entradas simuladas que se incluirán en el resultado. De forma predeterminada, todas las entradas simuladas se incluyen en los resultados. Puede excluir las entradas simuladas de las salidas. Se excluirán de los gráficos de tornado, diagramas de dispersión y resultados tabulares.

Rangos de límite para destinos continuos. Puede especificar el rango de valores válidos para uno o más destinos continuos. Los valores que están fuera del rango especificado se excluyen de todos los análisis de salida asociados con los destinos. Para establecer un límite inferior, seleccione **Inferior** en la columna Límite y especifique un valor en la columna Mínimo. Para establecer un límite superior, seleccione **Superior** en la columna Límite y especifique un valor en la columna Máximo. Para establecer un límite inferior y un límite superior, seleccione **Ambos** en la columna Límite y especifique los valores en las columnas Mínimo y Máximo.

Formatos de visualización. Puede definir el formato utilizado cuando se visualizan los valores de destinos y entradas (tanto entradas fijas como simuladas).

Guardar

Guardar el plan de esta simulación. Puede guardar las especificaciones actuales de su simulación en un archivo de plan de simulación. Los archivos de plan de simulación tienen la extensión *.splan*. Puede volver

a abrir el plan en el Generador de simulaciones, y también puede realizar modificaciones y ejecutar la simulación. Puede compartir el plan de simulación con otros usuarios, que pueden ejecutarlo en el cuadro de diálogo Ejecutar simulación. Los planes de simulación incluyen todas las especificaciones excepto las siguientes: ajustes de Funciones de densidad; Ajustes de resultados de gráficos y tablas; Opciones avanzadas de ajustes, Distribución empírica y Semilla aleatoria.

Guardar los datos simulados como un nuevo archivo de datos. Puede guardar entradas simuladas, entradas fijas y valores de destino pronosticados en un archivo de datos SPSS Statistics, un nuevo conjunto de datos en la sesión actual o un archivo Excel. Cada caso (o fila) del archivo de datos consta de los valores pronosticados de los objetivos junto con las entradas simuladas y las entradas fijas que generan los valores objetivo. Si se especifica el análisis de sensibilidad, cada iteración genera un conjunto contiguo de casos que se etiquetan con el número de iteración.

Cuadro de diálogo Ejecutar simulación

El cuadro de diálogo Ejecutar simulación está diseñado para usuarios que tienen un plan de simulación y que desean preferentemente ejecutar la simulación. También proporciona las características que necesita para ejecutar la simulación en condiciones diferentes. Permite ejecutar las siguientes tareas generales:

- Configurar o modificar los análisis de sensibilidad para investigar el efecto de variar el valor de una entrada fija o variar un parámetro de distribución de una entrada simulada.
- Reajustar distribuciones de probabilidad de entradas inciertas (y correlaciones entre esas entradas) a nuevos datos.
- Modificar la distribución de una entrada simulada.
- Personalizar los resultados.
- Ejecutar la simulación.

Pestaña Simulación

La pestaña Simulación permite especificar análisis de sensibilidad, reajustar distribuciones de probabilidad de entradas simuladas y correlaciones entre entradas simuladas en nuevos datos y modificar la distribución de probabilidad asociada con una entrada simulada.

La cuadrícula Entradas simuladas contiene una entrada para cada campo de entrada que se define en el plan de simulación. Cada entrada muestra el nombre y el tipo de distribución de probabilidad asociada con la entrada, junto con un gráfico de muestra de la curva de distribución asociada. Cada entrada también tiene un icono de estado asociado (un círculo de color con una marca de verificación) que es útil si está reajustando distribuciones a nuevos datos. Además, las entradas pueden incluir un icono de candado que indica que la entrada está bloqueada y no se puede modificar o reajustar a nuevos datos en el cuadro de diálogo Ejecutar simulación. Para modificar una entrada bloqueada, necesitará abrir el plan de simulación en el Generador de simulaciones.

Cada entrada puede ser simulada o fija. Las entradas simuladas son aquellas cuyos valores son inciertos y se generarán a partir de una distribución de probabilidad especificada. Las entradas fijas son aquellas cuyos valores se conocen y permanecen constantes en cada caso generado en la simulación. Para trabajar con una entrada concreta, seleccione la entrada en la cuadrícula Entradas simuladas.

Especificación de análisis de sensibilidad

El análisis de sensibilidad le permite investigar el efecto de los cambios sistemáticos en una entrada fija o en un parámetro de distribución para una entrada simulada generando un conjunto independiente de casos simulados -efectivamente, una simulación separada- para cada valor especificado. Para especificar los análisis de sensibilidad, seleccione una entrada fija o simulada y pulse en **Análisis de sensibilidad**. El análisis de sensibilidad está limitado a una única entrada fija o un parámetro de distribución único de una entrada simulada. Consulte el tema “Análisis de sensibilidad ” en la página 302 para obtener más información.

Reajuste de distribuciones a nuevos datos

Para reajustar automáticamente las distribuciones de probabilidad de entradas simuladas (y las correlaciones entre entradas simuladas) a los datos del conjunto de datos activo:

1. Compruebe que todas las entradas del modelo coinciden con el campo correcto del conjunto de datos activo. Todas las entradas simuladas se reajustan al campo del conjunto de datos activo especificado en la lista desplegable **Campo** con esa entrada. Puede identificar fácilmente las entradas no coincidentes buscando entradas con un icono de estado que incluya una marca de verificación con un signo de interrogación, tal y como se muestra a continuación.

2. Modifique cualquier campo necesario coincidente seleccionando **Ajustar a un campo en el conjunto de datos** y seleccionando el campo de la lista.
3. Pulse **Ajustar todas**.

En cada entrada ajustada, la distribución que mejor se ajusta a los datos se muestra junto con una representación de la distribución superpuesta en un histograma (o gráfico de barras) de los datos históricos. Si no se encuentra un ajuste aceptable, se utilizará la distribución empírica. En el caso de las entradas que se ajusten a la distribución empírica, solo verá un histograma de los datos históricos porque la distribución empírica está, de hecho, representada por ese histograma.

Nota: para ver una lista completa de iconos de estado, consulte el tema "Campos simulados" en la página 299.

Modificación de las distribuciones de probabilidad

Puede modificar la distribución de probabilidad de una entrada simulada y, opcionalmente, cambiar una entrada simulada a una entrada fija o viceversa.

1. Seleccione la entrada y seleccione **Ajustar la distribución manualmente**.
2. Seleccione el tipo de distribución y especifique los parámetros de distribución. Para cambiar una entrada simulada por una entrada fija, seleccione Fija en la lista desplegable **Tipo**.

Una vez haya introducido los parámetros de una distribución, el gráfico de muestra de la distribución (se muestra en la entrada) se actualizará reflejando sus cambios. Para obtener más información sobre la especificación manual de distribuciones de probabilidad, consulte el tema "Campos simulados" en la página 299.

Incluir valores perdidos del usuario de entradas categóricas al ajustar. Esto especifica si los valores perdidos del usuario de las entradas con una distribución categórica se tratan como válidos cuando se vuelven a ajustar los datos en el conjunto de datos activo. Los valores perdidos del sistema para todos los tipos de entrada se tratan siempre como no válidos. Todas las entradas deben tener valores válidos para que un caso se incluya en el ajuste de distribución y el cálculo de correlaciones.

Pestaña Resultado

La pestaña Resultado permite personalizar los resultados que genera la simulación.

Funciones de densidad. Las funciones de densidad son las medias principales de la prueba del conjunto de resultados de su simulación.

- **Función de densidad de probabilidad** La función de densidad de probabilidad muestra la distribución de los valores de destino, lo que permite determinar la probabilidad de que el destino esté dentro de una región concreta. En destinos con un conjunto fijo de resultados, como "servicio deficiente", "servicio correcto", "buen servicio" y "excelente servicio", se genera un gráfico de barras que muestra el porcentaje de casos que entran en cada categoría del destino.
- **Función de distribución acumulada.** La función de distribución acumulada muestra la probabilidad de que el valor del destino sea menor o igual que un valor especificado.

Gráficos de tornado. Los gráficos de tornado son gráficos de barra que muestran relaciones entre destinos y entradas simuladas utilizando una variedad de métricas.

- **Correlación del destino con la entrada.** Esta opción crea un gráfico de tornado de los coeficientes de correlación entre un destino especificado y cada una de sus entradas simuladas.
- **Contribución a la varianza.** Esta opción crea un gráfico de tornado que muestra la contribución a la varianza de un destino de cada una de sus entradas simuladas, lo que permite evaluar el grado en el que cada entrada contribuye a la incertidumbre general del destino.
- **Sensibilidad del destino para cambiar.** Esta opción crea un gráfico de tornado que muestra el efecto del destino del destino de modulación de cada entrada simulada más o menos una desviación estándar de la distribución asociada con la entrada.

Diagramas de dispersión de destinos frente a entradas. Esta opción genera diagramas de dispersión de destinos frente a entradas simuladas.

Diagramas de cajas de distribuciones de destino. Esta opción genera diagramas de cajas de las distribuciones de destino.

Tabla de cuartiles. Esta opción genera una tabla de cuartiles de las distribuciones de destino. Los cuartiles de una distribución son los percentiles 25, 50 y 75 de la distribución y dividen las observaciones en cuatro grupos de igual tamaño.

Correlaciones y tabla de contingencia para entradas. Esta opción muestra una tabla de coeficientes de correlación entre entradas simuladas. Una tabla de contingencia de asociaciones entre entradas con una distribución categórica se muestra cuando el plan de simulación especifica la generación de datos categóricos a partir de una tabla de contingencia.

Superponer resultados de destinos diferentes. Si el modelo predictivo que simula contiene múltiples destinos, puede especificar si los resultados de destinos diferentes se muestran en un único gráfico. Este ajuste se aplica a los gráficos de funciones de densidad de probabilidad, funciones de distribución acumuladas y diagramas de cajas. Por ejemplo, si selecciona esta opción, las funciones de densidad de probabilidad de todos los destinos se mostrarán en un gráfico único.

Guardar el plan de esta simulación. Puede guardar las modificaciones de su simulación en un archivo de plan de simulación. Los archivos de plan de simulación tienen la extensión *.splan*. Puede volver a abrir el plan en el cuadro de diálogo Ejecutar simulación o en el Generador de simulaciones. Los planes de simulación incluyen todas las especificaciones excepto los ajustes de resultados:

Guardar los datos simulados como un nuevo archivo de datos. Puede guardar entradas simuladas, entradas fijas y valores de destino pronosticados en un archivo de datos SPSS Statistics, un nuevo conjunto de datos en la sesión actual o un archivo Excel. Cada caso (o fila) del archivo de datos consta de los valores pronosticados de los objetivos junto con las entradas simuladas y las entradas fijas que generan los valores objetivo. Si se especifica el análisis de sensibilidad, cada iteración genera un conjunto contiguo de casos que se etiquetan con el número de iteración.

Si requiere más personalizaciones de un resultado que las disponibles aquí, considere ejecutar su simulación desde el Generador de simulaciones. Consulte el tema “Para ejecutar una simulación de un plan de simulación” en la página 295 para obtener más información.

Trabajo con resultados de gráficos de simulación

A partir de una simulación se generan diferentes gráficos que tienen características interactivas que permiten personalizar la vista. Las características interactivas están disponibles activando (pulsando dos veces) el objeto de gráfico en el Visor de resultados. Todos los gráficos de simulación son visualizaciones de tablero.

Gráficos de funciones de densidad de probabilidad de destinos continuos. Este gráfico tiene dos líneas de referencia verticales deslizantes que dividen el gráfico en regiones diferentes. La tabla bajo el gráfico muestra la probabilidad de que el destino esté en cada una de las regiones. Si se muestran múltiples funciones de densidad en el mismo gráfico, la tabla tiene una fila separada para las probabilidades asociadas con cada función de densidad. Cada una de las líneas de referencia tiene un deslizador (un triángulo invertido) que permite mover fácilmente la línea. Existe un número adicional de características disponibles pulsando en el botón **Opciones de gráfico**. En concreto, puede definir explícitamente las posiciones de los deslizadores, añadir líneas de referencia fijas y cambiar la vista del gráfico de una curva

continua a un histograma o viceversa. Consulte el tema “Opciones de gráfico ” en la página 310 para obtener más información.

Gráficos de funciones de distribución acumulada de destinos continuos. Este gráfico tiene las dos mismas líneas de referencia verticales móviles y la tabla asociada que se describe en el gráfico de funciones de densidad de probabilidad anterior. También proporciona acceso al cuadro de diálogo Opciones de gráfico, que permite definir explícitamente las posiciones de los deslizadores, añadir líneas de referencia fijas y especificar si la función de distribución acumulada se muestra como una función creciente (la opción predeterminada) o una función decreciente. Consulte el tema “Opciones de gráfico ” en la página 310 para obtener más información.

Gráficos de barras de destinos categóricos con iteraciones del análisis de sensibilidad. En los destinos categóricos con iteraciones de análisis de sensibilidad, los resultados de la categoría de destino pronosticada se muestran como un gráfico de barras agrupadas que incluye los resultados de todas las iteraciones. El gráfico incluye una lista desplegable que permite agrupar según la categoría o la iteración. En modelos de clúster de dos fases y modelos de clúster de K-medias, puede seleccionar agrupar según el número o iteración de clústeres.

Diagramas de cajas de múltiples destinos con iteraciones del análisis de sensibilidad. En modelos predictivos con múltiples destinos continuos e iteraciones del análisis de sensibilidad, si selecciona mostrar diagramas de cajas de todos los destinos en un único gráfico, se producirá un diagrama de cajas agrupado. El gráfico incluye una lista desplegable que permite agrupar según el destino o la iteración.

Opciones de gráfico

El cuadro de diálogo Opciones de gráfico permite personalizar la vista de los gráficos activados de funciones de densidad de probabilidad y funciones de distribución acumuladas generadas a partir de una simulación.

Vista. La lista desplegable **Ver** solo se aplica al gráfico de funciones de densidad de probabilidad. Permite cambiar la vista del gráfico de una curva continua a un histograma. Esta característica no está disponible si se muestran múltiples funciones de densidad en el mismo gráfico. En ese caso, las funciones de densidad solo se pueden visualizar como curvas continuas.

Orden. La lista desplegable **Ordenar** solo se aplica al gráfico de función de distribución acumulada. Especifica si la función de distribución acumulada se muestra como una función ascendente (la opción predeterminada) o una función descendente. Si se muestra como una función descendente, el valor de la función en un punto concreto del eje horizontal es la probabilidad de que el destino se encuentre a la derecha de ese punto.

Posiciones del deslizador. Puede definir explícitamente las posiciones de las líneas de referencia de los deslizadores introduciendo valores en los cuadros de texto Superior e Inferior. Puede eliminar la línea de la izquierda seleccionando **-Infinito**, definiendo la posición al infinito negativo, y puede eliminar la línea de la derecha seleccionando **Infinito**, definiendo la posición al infinito.

Líneas de referencia. Puede añadir varias líneas de referencia verticales fijas a funciones de densidad de probabilidad y funciones de distribución acumuladas. Cuando se muestran varias funciones en un único gráfico (debido a varios destinos o resultados de iteraciones del análisis de sensibilidad), puede especificar las funciones concretas a las que se aplican las líneas.

- **Sigmas.** Puede añadir líneas de referencia por encima o por debajo de un número especificado de desviaciones estándar desde la media de un objetivo.
- **Percentiles.** Ahora puede añadir líneas de referencia a uno o dos valores de los percentiles de la distribución de un objetivo introduciendo los valores en los cuadros de texto inferior y superior. Por ejemplo, un valor de 95 en el cuadro de texto superior representa el percentil 95, que es el valor por debajo del cual cae el 95% de las observaciones. Del mismo modo, un valor de 5 en el cuadro de texto inferior representa el percentil 5, que es el valor por debajo del cual cae el 5% de las observaciones.
- **Posiciones personalizadas.** Pueden añadirse líneas de referencia a los valores especificados en el eje horizontal.

Líneas de referencia de etiquetas. Esta opción controla si las etiquetas se aplican a las líneas de referencia seleccionadas.

Las líneas de referencia se eliminan anulando la selección asociada en el diálogo Opciones de gráfico y pulsando **Continuar**.

Modelado geoespacial

Las técnicas de modelado geoespacial están diseñadas para descubrir patrones de datos que incluyen un componente geoespacial (mapa). El sistema de modelado geoespacial proporciona métodos para analizar datos geoespaciales con y sin un componente de tiempo.

Busque asociaciones basadas en datos de evento y geoespaciales (reglas de asociación geoespacial)
Mediante reglas de asociación geoespacial, puede encontrar patrones en datos que se basan en las propiedades espaciales y no espaciales. Por ejemplo, podría identificar patrones en datos de delincuencia por ubicación y atributos demográficos. A partir de estos patrones, podrá generar reglas que predicen dónde es más probable que se vayan a producir determinados tipos de delitos.

Realice predicciones usando series temporales y datos geoespaciales (predicción espacio-temporal)
La predicción espacial temporal usa datos que contienen datos de ubicación, campos de entrada para la predicción (predictores), uno o más campos de hora y un campo objetivo. Cada ubicación tiene muchas filas en los datos que representan los valores de cada predictor y el objetivo en cada intervalo de tiempo.

Utilización del asistente de modelado geoespacial

1. En los menús seleccione:

   Analizar > **Modelado espacial y temporal** > **Modelado espacial**
2. Siga los pasos del asistente.

Ejemplos

Hay ejemplos detallados disponibles en el sistema de ayuda.

- Reglas de asociación geoespacial: **Ayuda** > **Temas** > **Estudios de casos** > **Statistics Base** > **Reglas de asociación espacial**
- Predicción temporal espacial: **Ayuda** > **Temas** > **Estudios de casos** > **Statistics Base** > **Predicción temporal espacial**

Selección de mapas

El modelado geoespacial puede utilizar uno o más orígenes de datos de mapas. Los orígenes de datos de mapas contienen información que define áreas geográficas y otras características geográficas como, por ejemplo, carreteras o ríos. Muchos orígenes de mapas también contienen datos demográficos u otros datos descriptivos y datos de evento como, por ejemplo, informes de delitos o tasas de desempleo. Puede utilizar un archivo de especificación de mapa definido previamente o definir especificaciones de mapas aquí y guardar estas especificaciones para su uso posterior.

Cargar especificación de mapa
Carga un archivo (.mplan) de especificación de mapa definido previamente. Los orígenes de datos de mapa que defina aquí se pueden guardar en un archivo de especificación de mapa. Para la predicción temporal espacial, si selecciona un archivo de especificación de mapa que identifica más de un mapa, se le solicitará que seleccione un mapa del archivo.

Añadir archivo de mapa
Añada un archivo de forma ESRI (.shp) o un archivo .zip que contenga un archivo de forma ESRI.

- Debe haber un archivo .dbf correspondiente en la misma ubicación que el archivo .shp, y dicho archivo debe tener el mismo nombre raíz que el archivo .shp.

- Si el archivo es un archivo .zip, los archivos .shp y .dbf deben tener el mismo nombre raíz que el archivo .zip.
- Si no hay ningún archivo (.prj) de proyección correspondiente, se le solicitará que seleccione un sistema de proyección.

Relación
Para las reglas de asociación geoespacial, esta columna define cómo están relacionados los eventos con las características en el mapa. Este valor no está disponible para la predicción temporal espacial.

Subir, Bajar
El orden de la capa de los elementos del mapa se determina mediante el orden en el cual aparecen en la lista. El primer mapa de la lista es la capa inferior.

Selección de un mapa

Para la predicción temporal espacial, si selecciona un archivo de especificación de mapa que identifica más de un mapa, se le solicitará que seleccione un mapa del archivo. La predicción temporal espacial no soporta varios mapas.

Relación geoespacial

Para las reglas de asociación geoespacial, el diálogo Relación geoespacial define cómo se relacionan los eventos con las características en el mapa.

- Este ajuste se aplica sólo a las reglas de asociación geoespacial.
- Este valor sólo afecta a orígenes de datos asociados con mapas especificados como datos de contexto en el paso de selección de orígenes de datos.

Relación

Cerrar
El evento se produce cerca de una área o punto especificados en el mapa.

Dentro de
El evento se produce dentro de un área especificada en el mapa.

Contiene
El área de evento contiene un objeto de contexto de mapa.

Intersecta
Las ubicaciones donde las líneas o regiones de los distintos mapas interseccionan entre sí.

Cruz
Para varios mapas, las ubicaciones donde las líneas (para carreteras, ríos, ferrocarriles) de distintas líneas se cruzan entre sí.

Norte de, Sur de, Este de, Oeste de
El evento se produce dentro de un área que está al norte, sur, este o oeste de un punto especificado en el mapa.

Establecer sistema de coordenadas

Si no hay ningún archivo de proyección (.prj) con el mapa o define dos campos en un origen de datos como un conjunto de coordenadas, debe definir el sistema de coordenadas.

Geográfico predeterminado (longitud y latitud)
El sistema de coordenadas es de longitud y latitud.

Cartesiano simple (X e Y)
El sistema de coordenadas es de coordenadas simples X e Y.

Utilizar un ID conocido (WKID)
Un "ID conocido" para proyecciones comunes.

Utilizar un nombre de sistema de coordenadas
El sistema de coordenadas se basa en la proyección denominada. El nombre se especifica entre paréntesis.

Definición de la proyección

Si no se puede determinar el sistema de proyección a partir de la información proporcionada con el mapa, debe especificar el sistema de proyección. La causa más habitual de esta condición es que falte un archivo de proyecto (.prj) asociado con el mapa o exista un archivo de proyección que no se pueda utilizar.

- **Una ciudad, región o país (Mercator)**
- **Un país grande, varios países o continentes (Winkel Tripel)**
- **Un área muy cercana al ecuador (Mercator)**
- **Un área cercana a uno de los polos (Stereographic)**

La proyección de Mercator es una proyección común utilizada en muchos mapas. Esta proyección trata el globo como un cilindro desplegado en una superficie plana. La proyección de Mercator distorsiona el tamaño y forma de los objetos grandes. Esta distorsión aumenta a medida que se aleja del ecuador y se acerca a los polos. Las proyecciones de Winkel Tripel y Stereographic realizan ajustes al hecho de que un mapa representa una parte de una esfera tridimensional que se visualiza en dos dimensiones.

Sistema de proyección y de coordenadas

Si selecciona más de una correlación y las correlaciones tienen distintos sistemas de proyección y coordenadas, debe seleccionar la correlación con el sistema de proyección que desea utilizar. Dicho sistema de proyección se utilizará para todas las correlaciones cuando se combinan juntos en el resultado.

Orígenes de datos

Un origen de datos puede ser un archivo dBase que se proporciona con el archivo de forma, un archivo de datos de IBM SPSS Estadística o un conjunto de datos abierto en la sesión actual.

Datos de contexto. Los datos de contexto identifican características en el mapa. Los datos de contexto también pueden contener campos que se pueden utilizar como entradas para el modelo. Para utilizar un archivo dBase (.dbf) de contexto asociado con un archivo de forma (.shp) de mapa, el archivo dBase de contexto debe estar en la misma ubicación que el archivo de forma y debe tener el mismo nombre raíz. Por ejemplo, si el archivo de forma es `geodata.shp`, el archivo dBase se debe llamar `geodata.dbf`

Datos de suceso. Los datos de evento contienen información sobre eventos que se producen como, por ejemplo, delitos o accidentes. Esta opción está disponible solo para reglas de asociación geoespacial.

Densidad de puntos. El intervalo de tiempo y los datos de coordenada para las estimaciones de la densidad de kernel. Esta opción sólo está disponible para la predicción temporal espacial.

Añadir. Abre un diálogo para añadir orígenes de datos. Un origen de datos puede ser un archivo dBase que se proporciona con el archivo de forma, un archivo de datos de IBM SPSS Estadística o un conjunto de datos abierto en la sesión actual.

Asociar. Abre un diálogo para especificar los identificadores (coordenadas o claves) utilizadas para asociar datos con mapas. Cada origen de datos debe contener uno o más identificadores que asocian los datos con el mapa. Los archivos dBase que vienen con un archivo de forma normalmente contienen un campo que se utiliza automáticamente como el identificador predeterminado. Para otros orígenes de datos, debe especificar los campos que se utilizan como identificadores.

Validar clave. Abre un diálogo para validar la coincidencia de claves entre el mapa y el origen de datos.

Reglas de asociación geoespacial

- Al menos, un origen de datos debe ser un origen de datos de evento.

- Todos los orígenes de datos de evento deben utilizar el mismo formato de identificadores de asociación de mapa: coordenadas o valores de clave.
- Si los orígenes de datos de evento se asocian a los mapas con valores de clave, todos los orígenes deben utilizar el mismo tipo de característica (por ejemplo, polígonos, puntos, líneas).

Predicción temporal espacial

- Debe haber un origen de datos de contexto.
- Si solo hay un origen de datos (un archivo de datos sin ningún mapa asociado), debe incluir valores de coordenadas.
- Si tiene dos orígenes de datos, un origen de datos debe ser datos de contexto y el otro origen de datos debe ser datos de densidad de puntos.
- No puede incluir más de dos orígenes de datos.

Añadir un origen de datos

Un origen de datos puede ser un archivo dBase que se proporciona con el archivo de forma y el archivo de contexto, un archivo de datos de IBM SPSS Estadística o un conjunto de datos abierto en la sesión actual.

Puede añadir el mismo origen de datos varias veces si desea utilizar una asociación espacial diferente con cada uno.

Datos y asociación de mapas

Cada origen de datos debe contener uno o más identificadores que asocian los datos con el mapa.

Coordenadas
: El origen de datos contiene campos que representan coordenadas cartesianas, seleccione los campos que representan dichas coordenadas X e Y. Para las reglas de asociación geoespacial, puede haber también una coordenada Z.

Valores de clave
: Los valores de clave en los campos del origen de datos corresponden a las claves de mapa seleccionadas. Por ejemplo, un mapa de regiones puede tener un identificador de nombre (clave de mapa) que etiqueta cada región. Dicho identificador se corresponde con un campo en los datos que también contienen los nombres de las regiones (clave de datos). Los campos se comparan con las claves de mapa basándose en el orden en el que se visualizan en las dos listas.

Validar claves

El diálogo Validar claves proporciona un resumen de coincidencias de registro entre la correlación y la fuente de datos, basándose en las claves de identificador seleccionadas. Si hay valores de clave de datos que no coinciden, puede hacer manualmente que coincidan con los valores de clave de correlación.

Reglas asociación geoespacial

Para las reglas de asociación geoespacial, tras definir mapas y orígenes de datos, los pasos restantes en el asistente son:

- Si hay varios orígenes de datos de evento, defina cómo se fusionan orígenes de datos de evento.
- Seleccione campos para utilizar como condiciones y predicciones en el análisis.

Si lo desea, tiene la posibilidad de:

- Seleccione distintas opciones de resultado.
- Guarde un archivo de modelo de puntuación.
- Cree campos nuevos para valores pronosticados y las reglas en los orígenes de datos utilizadas en el modelo.
- Personalice valores para crear regla de asociación.

- Personalice los valores de agrupación y agregación.

Definir campos de datos de eventos

Para las reglas de asociación geoespacial, si hay más de un origen de datos de evento, los orígenes de datos de evento se fusionan.

- De forma predeterminada, solo se incluyen los campos que son comunes para todos los orígenes de datos de evento.
- Puede visualizar una lista de campos comunes, campos para un origen de datos específico o campos de todos los orígenes de datos y seleccionar los campos que desea incluir.
- Para los campos comunes, los campos **Tipo** y **Medición** deben ser iguales para todos los orígenes de datos. Si hay conflictos, puede especificar el tipo y el nivel de medición para utilizar para cada campo común.

Seleccionar campos

La lista de campos disponibles incluye campos de los orígenes de datos de eventos y campos de los orígenes de datos de contexto.

- Puede controlar la lista de campos visualizados seleccionando un origen de datos en la lista **Orígenes de datos**.
- Debe seleccionar al menos dos campos. Al menos uno debe ser una condición y una debe ser, al menos, una predicción. Existen varias maneras de cumplir este requisito, incluyendo seleccionar dos campos para la lista **Ambas (condición y predicción)**.
- Las reglas de asociación predicen valores de los campos de predicción que se basan en valores de los campos de condición. Por ejemplo, en la regla "Si x=1 e y=2, entonces z=3", los valores de x e y son condiciones y el valor de z es la predicción.

Resultado

Tablas de reglas
: Cada tabla de reglas muestra las reglas superiores y los valores para la confianza, el soporte para regla, la elevación, el soporte de condición y la capacidad de despliegue. Cada tabla se clasifica por los valores del criterio seleccionado. Puede mostrar todas las reglas o el **Número** superior de reglas, basándose en el criterio seleccionado.

Nube de palabras clasificables
: Una lista de las reglas superiores, basándose en los valores del criterio seleccionado. El tamaño del texto indica la importancia relativa de la regla. El objeto de resultados interactivos contiene las reglas superiores para la confianza, el soporte de regla, la elevación, el soporte de condición y la capacidad de despliegue. El criterio seleccionado determina qué lista de reglas se visualiza de forma predeterminada. Puede seleccionar un criterio diferente de forma interactiva en el resultado. **Máx. de reglas que se va a mostrar** determina el número de reglas que se visualiza en el resultado.

Correlación
: Gráfico de barras interactivo y mapa de las reglas superiores, basándose en el criterio seleccionado. Cada objeto de resultados interactivos contiene las reglas superiores para la confianza, el soporte de regla, la elevación, el soporte de condición y la capacidad de despliegue. El criterio seleccionado determina qué lista de reglas se visualiza de forma predeterminada. Puede seleccionar un criterio diferente de forma interactiva en el resultado. **Máx. de reglas que se va a mostrar** determina el número de reglas que se visualiza en el resultado.

Tablas de información de modelo

Transformaciones de campo
: Describe las transformaciones que se aplican a campos utilizados en el análisis.

Resumen de registros.
: Número y porcentaje de registros incluidos y excluidos

Estadísticos de regla
Estadísticas de resumen para el soporte de condición, la confianza, el soporte de regla, la elevación y la capacidad de despliegue. Las estadísticas incluyen media, mínimo, máximo y desviación estándar.

Elementos más frecuentes.
Los elementos que aparecen con más frecuencia. Un elemento se incluye en una condición o una predicción en una regla. Por ejemplo, edad < 18 o sexo=mujer.

Campos más frecuentes.
Los campos que aparecen con más frecuencia en las reglas.

Entradas excluidas.
Los campos que se excluyen del análisis y la razón por la cual se ha excluido cada campo.

Criterio para tablas de reglas, nube de palabras y mapas

Confianza.
EL porcentaje de las predicciones de reglas correctas.

Soporte de regla.
El porcentaje de casos para los cuales la regla es verdadera. Por ejemplo, si la regla es "si x=1 e y=2, z=3," el soporte de la regla es el porcentaje real de casos en los datos para los cuales es correcto x=1, y=2 y z=3.

Elevación
La elevación es una medida que indica el grado de mejora que aplica la regla a la predicción en comparación con una predicción al azar. Es la proporción de predicciones correctas en la aparición global del valor pronosticado. El valor no puede ser mayor que 1. Por ejemplo, si el valor pronosticado se produce un 20 % del tiempo y la confianza en la predicción es del 80 %, el valor de elevación es 4.

Soporte de condición.
El porcentaje de casos para los cuales existe la condición de la regla. Por ejemplo, si la regla es "si x=1 e y=2, z=3," el soporte de condición es la proporción de casos en los cuales los datos son x=1 e y=2.

Capacidad de despliegue
El porcentaje de predicciones incorrectas cuando las condiciones son verdaderas. La capacidad de despliegue es igual a (1-confianza) multiplicado por el soporte de condición o soporte de condición menos soporte de regla.

Guardar

Guardar el mapa y los datos de contexto como especificación de mapa
Guarde las especificaciones del mapa en un archivo externo (.mplan). Puede cargar este archivo de especificación de mapa en el asistente para análisis posteriores. También puede utilizar el archivo de especificación de mapa con el comando SPATIAL ASSOCIATION RULES.

Copie cualquier archivo de mapa y datos en la especificación
Los datos de los archivos de forma de mapa, archivos de datos externos y conjuntos de datos utilizados en la especificación de mapa se guardan en el archivo de especificación de mapa.

Puntuación
Guarda los mejores valores de regla, los valores de confianza para las reglas y los valores de ID numérico para las reglas como campos nuevos en el origen de datos especificado.

Origen de datos para puntuar
El origen u orígenes de datos donde se crean los campos nuevos. Si el origen de datos no está abierto en la sesión actual, se abre en la sesión actual. Debe guardar explícitamente el archivo modificado para guardar los campos nuevos.

Valores de destino
Cree campos nuevos para los campos de objetivo seleccionado (predicción).

- Se crean dos campos nuevos para cada campo de objetivo: valor pronosticado y valor de confianza.

- Para los campos de objetivos continuos (escala), el valor pronosticado es una cadena que describe un rango de valores. Un valor de la forma "(value1, value2]" significa "mayor que el value1 y menor o igual que value2."

Número de mejores reglas
Cree campos nuevos para el número de mejores reglas especificado. Se crean tres campos nuevos para cada regla: valor de regla, valor de confianza y un valor de ID numérico para la regla.

Prefijo de nombre
Prefijo para utilizar para los nombres de campo nuevos.

Creación de reglas

Los parámetros de creación de reglas definen los criterios para las reglas de asociación generadas.

Elementos por regla
Número de valores de campo que se pueden incluir en las condiciones y predicciones de regla. EL número total de elemento no puede superar el 10. Por ejemplo, en la regla "si x=1 e y=2, z=3", hay dos elementos de condición y un elemento de predicción.

Número máximo de predicciones
El número máximo de valores de campo que se pueden producir en las predicciones para una regla.

Número máximo de condiciones
El número máximo de valores de campo que se pueden producir en las condiciones para una regla.

Excluir par
Excluye los pares de campos especificados de incluirse en la misma regla.

Criterio de regla

Confianza.
La confianza mínima que debe tener una regla para incluirse en el resultado. La confianza es el porcentaje de predicciones correctas.

Soporte de regla.
El soporte mínimo de regla que debe tener para que se incluya en el resultado. El valor representa el porcentaje de casos para los cuales la regla es verdadera en los datos observados. Por ejemplo, si la regla es "si x=1 e y=2, z=3," el soporte de la regla es el porcentaje real de casos en los datos para los cuales es correcto x=1, y=2 y z=3.

Soporte de condición.
Soporte de condición mínima que debe tener una regla para que se incluya en el resultado. EL valor representa el porcentaje de casos para los cuales existe la condición. Por ejemplo, si la regla es "si x=1 e y=2, z=3," el soporte de condición es el porcentaje de casos en los datos para los cuales x=1 e y=2.

Elevación
La elevación mínima que debe tener una regla para que se incluya en el resultado. La elevación es una medida que indica el grado de mejora que aplica la regla a la predicción con una predicción al azar. Es la proporción de predicciones correctas en la aparición global del valor pronosticado. Por ejemplo, si el valor pronosticado se produce un 20 % del tiempo y la confianza en la predicción es del 80 %, el valor de elevación es 4.

Tratar como igual
Identifica pares de campos, que se deben tratar como el mismo campo.

Agrupación y agregación

- La agregación es necesaria cuando hay más registros en los datos que características en el mapa. Por ejemplo, tiene registros de datos para países individuales pero tiene un mapa de estados.
- Puede especificar el método de medida de resumen de agregado para campos continuos y ordinales. Los campos nominales se agregan en función del valor modal.

Continuo
Para campos continuos (escala), la medida del resumen puede ser media, mediana o suma.

Ordinal
Para campos ordinales, la medida de resumen puede ser mediana, modo, mayor o menor.

Número de intervalos
Define el número máximo de agrupaciones para campos continuos (escala). Los campos continuos siempre se agrupan en rangos de valores. Por ejemplo, menor o igual que 5, mayor que 5 y menor o igual que 10, o mayor que 10.

Agregar el mapa
Aplique la agregación a datos y mapas.

Valores personalizados para campos específicos
Puede alterar temporalmente la medida de resumen predeterminada y el número de agrupaciones para campos específicos.

- Pulse el icono para abrir el diálogo **Selector de campos** y seleccione un campo para añadir a la lista.
- En la columna **Agregación**, seleccione una medida de resumen.
- Para campos continuos, pulse el botón en la columna **Agrupaciones** para especificar un número personalizado de agrupaciones para el campo en el diálogo **Agrupaciones**.

Predicción temporal espacial

Para la predicción temporal espacial, tras definir mapas y orígenes de datos, los pasos restantes del asistente son:

- Especifique el campo objetivo, los campos de tiempo y predictores opcionales.
- Defina los intervalos de tiempo o periodos cíclicos para los campos de tiempo.

Si lo desea, tiene la posibilidad de:

- Seleccione distintas opciones de resultado.
- Personalice parámetros de generación de modelos.
- Personalice valores de agregación.
- Guarde los valores pronosticados en un conjunto de datos en la sesión actual o en un archivo de datos de formato de IBM SPSS Estadística .

Seleccionar campos

La lista de campos disponibles incluye campos de los orígenes de datos seleccionados. Puede controlar la lista de campos visualizados seleccionando un origen de datos en la lista **Orígenes de datos**.

Destino
Es necesario un campo objetivo. El objetivo es el campo para el cual se han pronosticado valores.

- El campo objetivo debe ser un campo número continuo (escala).
- Si hay dos orígenes de datos, el objetivo es la estimación de densidad del kernel y la "densidad" se visualiza en el nombre del objetivo. No puede cambiar esta selección.

Predictores
Se pueden especificar uno o más campos de predictor. Este valor es opcional.

Campos de tiempo
Debe seleccionar uno o más campos que representan periodos de tiempo o seleccionar **Periodos cíclicos**.

- Si hay dos orígenes de datos, debe seleccionar campos de tiempo de ambos orígenes de datos. Ambos campos de tiempo deben representar el mismo intervalo.
- Para los periodos cíclicos, debe especificar los campos que definen ciclos de periodicidad en el **panel Intervalos de tiempo** del asistente.

Intervalos de tiempo

Las opciones de este panel se basan en la selección de **Campos de hora** o **Periodo cíclico** en el paso de selección de campos.

Campos de hora

Campos de hora seleccionados. Si selecciona uno o más campos de hora en el paso para seleccionar campos, estos campos se visualizan en esta lista.

Intervalo de tiempo. Seleccione el intervalo de tiempo apropiado en la lista. En función del intervalo de tiempo, también puede especificar otros valores como, por ejemplo, el intervalo entre observaciones (incremento) y el valor de inicio. Este intervalo de tiempo se utiliza para todos los campos de hora seleccionados.

- El procedimiento presupone que todos los casos (registros) representan intervalos espaciados de forma uniforme.
- Basándose en el intervalo de tiempo seleccionado, el procedimiento puede detectar observaciones que faltan o varias observaciones en el mismo intervalo de tiempo que se deben agregar de forma conjunta. Por ejemplo, si el intervalo de tiempo es días y la fecha 2014-10-27 viene seguida por 2014-10-29, hay una observación que falta para el 2014-10-28. Si el intervalo de tiempo es mes, varias fechas del mismo mes se agregan juntas.
- Para algunos intervalos de tiempo, el valor adicional puede definir saltos en los intervalos normales con un espacio uniforme. Por ejemplo, si el intervalo de tiempo es días, pero solo son válidos los fines de semana, puede especificar que hay cinco días en una semana y que la semana empieza el lunes.
- Si el campo de hora seleccionado no es un campo de formato de fecha ni de hora, el intervalo de tiempo se define automáticamente en **Periodos** y no se puede cambiar.

Campos de ciclo

Si seleccione el **Periodo cíclico** en el paso para seleccionar campos, debe especificar los campos que definen los periodos cíclicos. Un periodo cíclico identifica una variación cíclica repetitiva como, por ejemplo, el número de meses de un año o el número de días de una semana.

- Puede especificar hasta tres campos que definen periodos cíclicos.
- El primer campo de ciclo representa el nivel más alto del ciclo. Por ejemplo, si hay una variación cíclica por año, trimestre y mes, el campo que representa el año es el primer campo de ciclo.
- La longitud del ciclo para el primer y segundo campo del ciclo es la periodicidad en el nivel siguiente. Por ejemplo, si los campos de ciclo son año, trimestre y mes, la longitud del primer ciclo es 4 y la longitud del segundo ciclo es 3.
- El valor de inicio para los campos del segundo y el tercer ciclo es el primer valor de cada uno de estos periodos cíclicos.
- Los valores de longitud del ciclo y de inicio deben ser enteros positivos.

Agregación

- Si selecciona algún **Predictor** en el paso para seleccionar campos, puede seleccionar el método de resumen de agregación para los predictores.
- La agregación es necesaria cuando hay más de un registro en un intervalo de tiempo definido. Por ejemplo, si el intervalo de tiempo es Mes, varias fechas del mismo mes se agregan juntas.
- Puede especificar el método de medida de resumen de agregación para los campos continuos y ordinales. Los campos nominales se agregan en función del valor modal.

Continuo
Para campos continuos (escala), la medida del resumen puede ser media, mediana o suma.

Ordinal
Para campos ordinales, la medida de resumen puede ser mediana, modo, mayor o menor.

Valores personalizados para campos específicos
Puede alterar temporalmente la medida de resumen de agregación predeterminada para predictores específicos.

- Pulse el icono para abrir el diálogo **Selector de campos** y seleccione un campo para añadir a la lista.
- En la columna **Agregación**, seleccione una medida de resumen.

Resultado

Mapas

Valores de destino
Mapa de valores para el campo objetivo seleccionado.

Correlación
Mapa de correlaciones.

Clústeres
Mapa que resalta los clústeres de ubicaciones que son similares entre sí. Hay mapas de clústeres disponibles sólo para modelos empíricos.

Umbral de similitud de ubicación.
La similitud necesaria para crear clústeres. El valor debe ser un número mayor que cero y menor que 1.

Especifique el número máximo de clústeres
El número máximo de clústeres para visualizar.

Tablas de evaluación de modelos

Especificaciones de modelos
Resumen de especificaciones utilizadas para ejecutar el análisis, incluidos campos objetivo, de entrada y ubicación.

Resumen de información temporal.
Identifica los campos de tiempo y los intervalos de tiempo utilizados en el modelo.

Pruebas de efectos en estructura media
El resultado incluye el valor de estadístico de prueba, grados de libertad y el nivel de significación para el modelo y cada efecto.

Estructura media de los coeficientes de modelo
El resultado incluye el valor coeficiente, error estándar, valor de estadísticas de prueba, nivel de significación e intervalos de confianza para cada término del modelo.

Coeficientes autorregresivos
El resultado incluye el valor de coeficiente, error estándar, valor de estadísticas de prueba, nivel de significación e intervalos de confianza para cada retardo.

Pruebas de covarianza espacial.
Para los modelos paramétricos basados en variograma, se muestran los resultados de la prueba de bondad de ajuste para la estructura de covarianza espacial. Los resultados de la prueba pueden determinar si se debe modelar la estructura de covarianza espacial de forma paramétrica o utilizar un modelo no paramétrico.

Covarianza espacial paramétrica.
Para los modelos paramétricos basados en variograma, se muestran las estimaciones de parámetro para la covarianza espacial paramétrica.

Opciones de modelo

Valores del modelo

Incluir automáticamente una intersección
Incluir la interceptación en el modelo.

Retardo máximo de autorregresión
El retardo máximo de autorregresión. El valor debe ser un entero entre 1 y 5.

Covarianza espacial
Especifica el método de estimación para la covarianza espacial.

Paramétrico
EL método de estimación es paramétrico. El método puede ser **Gauss**, **Exponencial** o **Potencia exponencial**. Para la potencia exponencial, puede especificar el valor **Potencia**.

No paramétrico
El método de estimación no es paramétrico.

Guardar

Guardar el mapa y los datos de contexto como especificación de mapa
Guarde las especificaciones del mapa en un archivo externo (.mplan). Puede cargar este archivo de especificación de mapa en el asistente para un análisis posterior. También puede utilizar el archivo de especificación de mapa con el comando `SPATIAL TEMPORAL PREDICTION`.

Copie cualquier archivo de mapa y datos en la especificación
Los datos de archivos de forma de mapa, archivos de datos externos y conjuntos de datos utilizados en la especificación del mapa se guardan en el archivo de especificación de mapa.

Puntuación
Guarda valores pronosticados, varianza y los límites de confianza superior e inferior para el campo objetivo en el archivo de datos seleccionado.

- Puede guardar valores pronosticados en un conjunto de datos abierto en la sesión actual o en archivo de datos de formato de IBM SPSS Estadística .
- El archivo de datos no puede ser un origen de datos utilizado en el modelo.
- El archivo de datos debe contener todos los campos de tiempo y predictores utilizados en el modelo.
- Los valores de tiempo deben ser mayores que los valores de tiempo utilizados en el modelo.

Avanzado

Número máximo de casos con valores perdidos (%)
El porcentaje máximo de casos con valores perdidos.

Nivel de significación
El nivel de significación para determinar si un modelo paramétrico basado en variograma es apropiado. El valor debe ser mayor que 0 y menor que 1. El valor predeterminado es 0,05. El nivel de significación se utiliza en la prueba de bondad de ajuste para la estructura de covarianza espacial. La estadística de bondad de ajuste se utiliza para determinar un modelo paramétrico o no paramétrico.

Factor de incertidumbre (%)
El factor de incertidumbre es un valor de porcentaje que representa el crecimiento de la incertidumbre al realizar predicciones de futuro. Los límites superior e inferior de la incertidumbre de predicción aumentan en el porcentaje especificado para cada paso en el futuro.

Finalizar

En el último paso del asistente de modelado geoespacial, puede ejecutar el modelo o pegar la sintaxis del comando generada en una ventana de sintaxis. Puede modificar y guardar la sintaxis generada para su uso posterior.

Avisos

Esta información se ha desarrollado para productos y servicios ofrecidos en Estados Unidos. Este material puede estar disponible en otros idiomas en IBM. Sin embargo, deberá poseer una copia del producto o de la versión del producto en ese idioma para poder acceder a él.

Es posible que IBM no ofrezca los productos, servicios o características que se tratan en este documento en otros países. Consulte al representante local de IBM para obtener información sobre los productos y servicios que actualmente pueden adquirirse en su zona. Las referencias a programas, productos o servicios de IBM no pretenden establecer ni implicar que sólo puedan utilizarse dichos productos, programas o servicios de IBM. En su lugar, se puede utilizar cualquier producto, programa o servicio funcionalmente equivalente que no infrinja ninguno de los derechos de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patente en tramitación que cubran la materia descrita en este documento. La posesión de este documento no le otorga ninguna licencia sobre dichas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
EE.UU.

Para consultas sobre licencias relacionadas con información de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus consultas, por escrito, a:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokio 103-8510, Japón

INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍAS DE NINGUNA CLASE, NI EXPLÍCITAS NI IMPLÍCITAS, QUE INCLUYEN, PERO NO SE LIMITAN A, LAS GARANTÍAS IMPLÍCITAS DE NO VULNERACIÓN, MERCANTIBILIDAD O ADECUACIÓN A UN FIN DETERMINADO. Algunas jurisdicciones no permiten la renuncia a las garantías explícitas o implícitas en ciertas transacciones; por lo tanto, es posible que esta declaración no sea aplicable en su caso.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Periódicamente se efectúan cambios en la información aquí contenida; estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede realizar mejoras y/o cambios en los productos y/o programas descritos en esta publicación en cualquier momento sin previo aviso.

Las referencias hechas en esta publicación a sitios web que no son de IBM se proporcionan sólo para la comodidad del usuario y no constituyen de modo alguno un aval de esos sitios web. Los materiales de estos sitios web no forman parte de los materiales para este producto IBM, por lo que la utilización de dichos sitios web es a cuenta y riesgo del usuario.

IBM puede utilizar o distribuir la información que se le proporcione del modo que considere adecuado sin incurrir por ello en ninguna obligación con el remitente.

Los tenedores de licencias de este programa que deseen obtener información acerca de éste con el fin de permitir: (i) el intercambio de información entre programas creados independientemente y otros programas (incluido el presente) y (ii) la utilización mutua de la información que se ha intercambiado, deben ponerse en contacto con:

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
EE.UU.

Dicha información puede estar disponible, sujeta a los términos y condiciones correspondientes, incluyendo, en algunos casos, el pago de una tarifa.

IBM proporciona el programa bajo licencia descrito en este documento y todo el material bajo licencia disponible para el mismo según los términos del acuerdo de cliente de IBM, el acuerdo internacional de licencia de programa de IBM o cualquier otro acuerdo equivalente entre las partes.

Los ejemplos de datos de rendimiento y de clientes citados se presentan solamente a efectos ilustrativos. Los resultados reales de rendimiento pueden variar en función de las configuraciones específicas y condiciones de operación.

La información acerca de productos que no son de IBM se ha obtenido de los proveedores de dichos productos, sus publicaciones anunciadas u otras fuentes disponibles públicamente. IBM no ha probado esos productos y no puede confirmar la exactitud del rendimiento, la compatibilidad ni ninguna otra afirmación relacionada con productos no IBM. Las preguntas relacionadas con las prestaciones de los productos que no son de IBM deben dirigirse a los proveedores de esos productos.

Las declaraciones sobre el futuro rumbo o intención de IBM están sujetas a cambio o retirada sin previo aviso y representan únicamente metas y objetivos.

Esta información contiene ejemplos de datos e informes utilizados en operaciones empresariales cotidianas. Para ilustrarlos de la forma más completa posible, los ejemplos incluyen nombres de personas, empresas, marcas y productos. Todos estos nombres son ficticios y cualquier parecido con los nombres de personas o empresas reales es pura coincidencia.

LICENCIA DE DERECHOS DE AUTOR:

Esta información contiene programas de aplicación de muestra en lenguaje fuente, que ilustran técnicas de programación en distintas plataformas operativas. Puede copiar, modificar y distribuir estos programas de ejemplo en cualquier formato sin abonar ninguna cantidad a IBM, con el fin de desarrollar, utilizar, comercializar o distribuir programas de aplicaciones de acuerdo con la interfaz de programación de aplicaciones correspondiente a la plataforma operativa para la que se han escrito los programas de ejemplo. Estos ejemplos no se han probado exhaustivamente bajo todas las condiciones. Por lo tanto, IBM no puede garantizar ni dar por supuesta la fiabilidad, la capacidad de servicio ni la funcionalidad de estos programas. Los programas de ejemplo se proporcionan "TAL CUAL" sin garantías de ningún tipo. IBM no será responsable de ningún daño resultante del uso de los programas de ejemplo.

Cada copia, parcial o completa, de estos programas de ejemplo, o cualquier trabajo obtenido a partir de los mismos, debe incluir el siguiente aviso de copyright:

© Copyright IBM Corp. 2021. Las partes de este código se derivan de IBM Corp. Programas de ejemplo

© Copyright IBM Corp. 1989 - 2021. Reservados todos los derechos.

Marcas comerciales

IBM, el logotipo de IBM e ibm.com son marcas comerciales o marcas registradas de International Business Machines Corp., registradas en numerosas jurisdicciones de todo el mundo. Otros nombres de productos y servicios pueden ser marcas registradas de IBM o de otras empresas. En la web hay disponible una lista actualizada de las marcas registradas de IBM en "Copyright and trademark information" en www.ibm.com/legal/copytrade.shtml.

Adobe, el logotipo Adobe, PostScript y el logotipo PostScript son marcas registradas o marcas comerciales de Adobe Systems Incorporated en Estados Unidos y/o otros países.

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en Estados Unidos y otros países.

Linux es una marca registrada de Linus Torvalds en los Estados Unidos o en otros países.

Microsoft, Windows, Windows NT, y el logotipo de Windows son marcas comerciales de Microsoft Corporation en Estados Unidos, otros países o ambos.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Java y todas las marcas registradas y logotipos basados en Java son marcas registradas de Oracle o sus filiales.

Índice

Caracteres Especiales

índice de concentración
    en Estadísticos de la razón 282
último
    en Cubos OLAP 105
    en Medias 103
    en Resumir 101

A

Acuerdo entre evaluadores 269
Agresti-Caffo
    en proporciones de muestras independientes 115
Agresti-Coull
    en proporciones de una muestra 109
Agresti-Min
    en proporciones de muestras emparejadas 112
ajuste de distribución
    en simulación 299
ajuste de distribución automático
    en simulación 299
alfa de Cronbach
    en Análisis de fiabilidad 268, 269
análisis de clústeres
    Análisis de clústeres de K-medias 234
    Análisis de clústeres jerárquico 232
    eficacia 235
Análisis de clústeres bietápicos
    almacenamiento en el archivo de trabajo 227
    almacenamiento en un archivo externo 227
    estadísticos 227
    opciones 226
Análisis de clústeres de K-medias
    Características adicionales del comando 237
    clúster de pertenencia 236
    criterios de convergencia 236
    distancias entre clústeres 236
    eficacia 235
    ejemplos 234
    estadísticos 234, 236
    guardar información de clústeres 236
    iteraciones 236
    métodos 234
    valores perdidos 236
    visión general 234
Análisis de clústeres jerárquico
    Características adicionales del comando 234
    casos de clúster 232
    clúster de pertenencia 234
    dendrogramas 234
    diagramas de témpanos 234
    ejemplo 232
    estadísticos 232, 234
    guardar variables nuevas 234
    historial de conglomeración 234
    matrices de distancias 234
Análisis de clústeres jerárquico *(continuación)*
    medidas de distancia 233
    medidas de similaridad 233
    métodos de agrupación en clústeres 233
    orientación de los gráficos 234
    transformar medidas 233
    transformar valores 233
    variables de clúster 232
análisis de componentes principales 220, 221
Análisis de fiabilidad
    Características adicionales del comando 271
    coeficiente de correlación intraclase 269
    correlaciones y covarianzas entre elementos 269
    descriptivos 269
    ejemplo 268
    estadísticos 268, 269
    Kuder-Richardson 20 269
    Prueba de aditividad de Tukey 269
    T 2 de Hotelling 269
    tabla de ANOVA 269
análisis de la varianza
    en ANOVA de un factor 121
    en Estimación curvilínea 204
    en Medias 103
    en Regresión lineal 175
Análisis de potencia
    estadísticos 1
análisis de respuestas múltiples
    Frecuencias de respuestas múltiples 260
    tabla cruzada 261
    Tablas cruzadas de respuestas múltiples 261
    tablas de frecuencias 260
análisis de sensibilidad
    en simulación 302
análisis de series temporales
    predicción 205
    predicción de casos 205
Análisis de vecinos más cercanos
    guardar variables 212
    opciones 212
    particiones 211
    salida 212
    selección de características 210
    vecinos 210
    vista de modelo 213
Análisis discriminante
    almacenamiento de variables de clasificación 219
    Características adicionales del comando 219
    coeficientes de función 217
    criterios 218
    definición de rangos 217
    Distancia de Mahalanobis 218
    ejemplo 216
    estadísticos 216, 217
    estadísticos descriptivos 217
    exportación de información del modelo 219
    gráficos 218

Análisis discriminante *(continuación)*
Lambda de Wilks 218
matrices 217
matriz de covarianzas 218
métodos de inclusión por pasos 216
métodos discriminantes 218
opciones de representación 218
probabilidades previas 218
selección de casos 217
V de Rao 218
valores perdidos 218
variables de agrupación 216
variables independientes 216
Análisis factorial
Características adicionales del comando 223
convergencia 221, 222
descriptivos 221
ejemplo 220
estadísticos 220, 221
formato de presentación de los coeficientes 223
gráficos de cargas 222
métodos de extracción 221
métodos de rotación 222
puntuaciones factoriales 223
selección de casos 221
valores perdidos 223
visión general 220
análisis hipotético
en simulación 302
Análisis PROXMAP 143, 146
Análisis ROC
estadísticos y gráficos 290, 291
ANOVA
en ANOVA de un factor 121
en GLM Univariante 126
en Medias 103
en modelos lineales 169
modelo 128
ANOVA de un factor
Características adicionales del comando 126
comparaciones múltiples 123
contrastes 123
contrastes polinómicos 123
estadísticos 125
opciones 125
pruebas post hoc 123
valores perdidos 125
variables del factor 121
Anscombe
en proporciones de una muestra 109
asignación de memoria
en Análisis de clústeres en dos fases 226
asimetría
en Cubos OLAP 105
en Descriptivos 91
en el Informe de estadísticos en filas 264
en el Resúmenes de informes en columnas 266
en Explorar 93
en Frecuencias 88
en Medias 103
en Resumir 101
asociación lineal por lineal
en Tablas cruzadas 97
autovalores
autovalores *(continuación)*
en Análisis factorial 221
en Regresión lineal 175

B

bagging
en modelos lineales 165
Binomial ajustado mid-p
en proporciones de muestras emparejadas 112
Binomial exacto
en proporciones de muestras emparejadas 112
Bland-Altman
análisis de acuerdos 138, 139
bondad de ajuste
en regresión ordinal 178
Bonett-Price
en proporciones de muestras emparejadas 112
Bonferroni
en ANOVA de un factor 123
en MLG 132
boosting
en modelos lineales 165
Brown-Li-Jeffreys
en proporciones de muestras independientes 115

C

C de Dunnett
en ANOVA de un factor 123
en MLG 132
capas
en Tablas cruzadas 96
CARTOGRAFÍA DE PROXIMIDAD 143, 146
categoría de referencia
en MLG 129, 130
CCI. Consulte el coeficiente de correlación intraclase 269
chi-cuadrado
asociación lineal por lineal 97
corrección por continuidad de Yates 97
en Tablas cruzadas 97
estadísticos 249
opciones 249
para la independencia 97
Pearson 97
prueba exacta de Fisher 97
prueba para una muestra 248
rango esperado 248
razón de verosimilitud 97
valores esperados 248
valores perdidos 249
Chi-cuadrado
Prueba de hipótesis estadística 139–141
chi-cuadrado de la razón de verosimilitud
en regresión ordinal 178
en Tablas cruzadas 97
Chi-cuadrado de Pearson
en regresión ordinal 178
en Tablas cruzadas 97
ChiSquare 139–141
clasificación
en Análisis ROC 289
en Curva ROC... 292

Clopper-Pearson (exacto)
en proporciones de una muestra 109
clústeres
presentación de clústeres 228
presentación global 228
selección de procedimientos 224
coeficiente alfa
en Análisis de fiabilidad 268, 269
Coeficiente de concordancia de Kendall (W)
Pruebas no paramétricas de muestras relacionadas 245
coeficiente de contingencia
en Tablas cruzadas 97
coeficiente de correlación de los rangos
en Correlaciones bivariadas 141
coeficiente de correlación de Spearman
en Correlaciones bivariadas 141
en Tablas cruzadas 97
coeficiente de correlación intraclase (CCI)
en Análisis de fiabilidad 269
coeficiente de correlación r
en Correlaciones bivariadas 141
en Tablas cruzadas 97
coeficiente de dispersión (CDD)
en Estadísticos de la razón 282
coeficiente de incertidumbre
en Tablas cruzadas 97
coeficiente de variación (CDV)
en Estadísticos de la razón 282
coeficientes beta
en Regresión lineal 175
coeficientes de regresión
en Regresión lineal 175
columna total
en informes 266
comparación de grupos
en Cubos OLAP 107
comparación de métodos
en Bland Altman 138
comparación de variables
en Cubos OLAP 107
comparaciones múltiples
en ANOVA de un factor 123
comparaciones múltiples post hoc 123
conjuntos
en modelos lineales 167
conjuntos de respuestas múltiples
Libro de códigos 84
contrastes
en ANOVA de un factor 123
en MLG 129, 130
contrastes de desviación
en MLG 129, 130
contrastes de diferencia
en MLG 129, 130
Contrastes de Helmert
en MLG 129, 130
contrastes de linealidad
en Medias 103
contrastes polinómicos
en ANOVA de un factor 123
en MLG 129, 130
contrastes repetidos
en MLG 129, 130
contrastes simples
contrastes simples *(continuación)*
en MLG 129, 130
control de página
en el informe de estadísticos en columnas 267
en informes de estadísticos en filas 264
convergencia
en Análisis de clústeres de K-medias 236
en Análisis factorial 221, 222
corrección por continuidad de Yates
en Tablas cruzadas 97
Correlación de distancias 161
Correlación de Pearson
en Correlaciones bivariadas 141
en Tablas cruzadas 97
correlaciones
de orden cero 159
en Correlaciones bivariadas 141
en Correlaciones parciales 158
en simulación 303
en Tablas cruzadas 97
Correlaciones bivariadas
Características adicionales del comando 143
coeficientes de correlación 141
estadísticos 142
intervalo de confianza 141
intervalos de confianza 143
nivel de significación 141
opciones 142
valores perdidos 142
correlaciones de orden cero
en Correlaciones parciales 159
Correlaciones parciales
Características adicionales del comando 160
correlaciones de orden cero 159
en Regresión lineal 175
estadísticos 159
opciones 159
valores perdidos 159
crear términos 128, 179, 180
Criterio de información de Akaike
en modelos lineales 166
criterio de prevención sobreajustado
en modelos lineales 166
criterios
en Bland Altman 139
criterios de información
en modelos lineales 166
criterios de selección
en Regresión lineal 175
cuadro de diálogo Contraste
Meta-análisis binario 60
Meta-análisis continuo 44
Tamaño de efecto de meta-análisis binario 69
Tamaño de efecto de meta-análisis continuo 52
cuartiles
en Frecuencias 88
Cubos OLAP
estadísticos 105
títulos 107
curtosis
en Cubos OLAP 105
en Descriptivos 91
en el Informe de estadísticos en filas 264
en el Resúmenes de informes en columnas 266

curtosis *(continuación)*
en Explorar 93
en Frecuencias 88
en Medias 103
en Resumir 101
Curva ROC
estadísticos y gráficos 293

D

d
en Tablas cruzadas 97
d de Somers
en Tablas cruzadas 97
de seguridad
columnas totales 266
comparación de columnas 266
división de valores de las columnas 266
informe de estadísticos en columnas 265
informes de estadísticos en filas 263
multiplicación de valores de las columnas 266
totales compuestos 266
Definir conjuntos de respuestas múltiples
categorías 259
dicotomías 259
etiquetas del conjunto 259
nombres del conjunto 259
dendrogramas
en Análisis de clústeres jerárquico 234
descomposición jerárquica 128
Descriptivos
almacenamiento de puntuaciones Z 90
Características adicionales del comando 91
estadísticos 91
orden de visualización 91
desviación absoluta promedio (DAP)
en Estadísticos de la razón 282
desviación estándar
en Cubos OLAP 105
en Descriptivos 91
en el Informe de estadísticos en filas 264
en el Resúmenes de informes en columnas 266
en Estadísticos de la razón 282
en Explorar 93
en Frecuencias 88
en Medias 103
en Resumir 101
DfBeta
en Regresión lineal 173
DfFit
en Regresión lineal 173
diagrama de dispersión
en simulación 305
diagramas de caja
comparación de variables 94
comparar niveles del factor 94
en Explorar 94
en simulación 305
diagramas de dispersión
en Regresión lineal 173
diagramas de témpanos
en Análisis de clústeres jerárquico 234
diálogo Análisis
Meta-análisis binario 58
diálogo Análisis *(continuación)*
Meta-análisis continuo 42
Tamaño de efecto de meta-análisis binario 67
Tamaño de efecto de meta-análisis continuo 50
diálogo Criterios
Meta-análisis binario 57
Meta-análisis continuo 41, 49
Metaanálisis Regresión 75
Tamaño de efecto de meta-análisis binario 66
diálogo Gráfico
Meta-análisis binario 65
Meta-análisis continuo 48
Metaanálisis Regresión 77
Tamaño de efecto de meta-análisis binario 73
Tamaño de efecto de meta-análisis continuo 56
diálogo Guardar
Meta-análisis binario 64
Meta-análisis continuo 47
Metaanálisis Regresión 77
Tamaño de efecto de meta-análisis binario 73
Tamaño de efecto de meta-análisis continuo 55
diálogo Imprimir
Meta-análisis binario 63
Meta-análisis continuo 47
Metaanálisis Regresión 76
Tamaño de efecto de meta-análisis binario 72
Tamaño de efecto de meta-análisis continuo 54
diálogo Inferencia
Meta-análisis binario 59
Meta-análisis continuo 43
Metaanálisis Regresión 75
Tamaño de efecto de meta-análisis binario 68
Tamaño de efecto de meta-análisis continuo 51
diálogo Recortar y rellenar
Meta-análisis binario 62
Meta-análisis continuo 45
Tamaño de efecto de meta-análisis binario 70
Tamaño de efecto de meta-análisis continuo 53
diálogo Sesgo
Meta-análisis binario 61
Meta-análisis continuo 45
Tamaño de efecto de meta-análisis binario 70
Tamaño de efecto de meta-análisis continuo 52
diccionario
Libro de códigos 84
diferencia honestamente significativa de Tukey
en ANOVA de un factor 123
en MLG 132
diferencia menos significativa
en ANOVA de un factor 123
en MLG 132
diferencial relativo al precio (DRP)
en Estadísticos de la razón 282
diferencias entre grupos
en Cubos OLAP 107
diferencias entre variables
en Cubos OLAP 107
diseños para las pruebas de heterocedasticidad
en MLG 137
distancia chi-cuadrado
en Distancias 160
distancia de bloques
en Distancias 160
distancia de bloques de ciudad

distancia de bloques de ciudad *(continuación)*
en Análisis de vecinos más próximos 210
distancia de Chebychev
en Distancias 160
Distancia de Cook
en MLG 135
en Regresión lineal 173
Distancia de Mahalanobis
en Análisis discriminante 218
en Regresión lineal 173
Distancia de Manhattan
en Análisis de vecinos más próximos 210
distancia de Minkowski
en Distancias 160
Distancia euclídea
en Análisis de vecinos más próximos 210
en Distancias 160
distancia euclídea al cuadrado
en Distancias 160
Distancias
cálculo de distancias entre casos 160
cálculo de distancias entre variables 160
Características adicionales del comando 161
ejemplo 160
estadísticos 160
medidas de disimilaridad 160
medidas de similaridad 161
transformar medidas 160, 161
transformar valores 160, 161
distancias de vecinos más próximos
en Análisis de vecinos más próximos 215
división
división entre columnas del informe 266
DMS de Fisher
en MLG 132

E

eliminación hacia atrás
en Regresión lineal 172
enlace
en regresión ordinal 177
error estándar
en Análisis ROC 290, 291
en Curva ROC... 293
en Descriptivos 91
en Explorar 93
en Frecuencias 88
en MLG 135
error estándar de la asimetría
en Cubos OLAP 105
en Medias 103
en Resumir 101
error estándar de la curtosis
en Cubos OLAP 105
en Medias 103
en Resumir 101
error estándar de la media
en Cubos OLAP 105
en Medias 103
en Resumir 101
escala
en Análisis de fiabilidad 268
en Escalamiento multidimensional 274
escala *(continuación)*
en Kappa ponderada 272
Escalamiento multidimensional
Características adicionales del comando 276
condicionalidad 276
crear matrices de distancias 275
criterios 276
definir la forma de los datos 275
dimensiones 276
ejemplo 274
estadísticos 274
medidas de distancia 275
modelos de escalamiento 276
niveles de medición 276
opciones de representación 276
transformar valores 275
estadístico de Brown-Forsythe
en ANOVA de un factor 125
estadístico de Cochran
en Tablas cruzadas 97
estadístico de Mantel-Haenszel
en Tablas cruzadas 97
estadístico de Welch
en ANOVA de un factor 125
estadístico Durbin-Watson
en Regresión lineal 175
estadístico F
en modelos lineales 166
estadístico R
en Medias 103
en Regresión lineal 175
Estadísticos de la razón
estadísticos 282
estadísticos de proporciones de columna
en Tablas cruzadas 99
estadísticos descriptivos
en Análisis de clústeres en dos fases 227
en Descriptivos 90
en Estadísticos de la razón 282
en Explorar 93
en Frecuencias 88
en Resumir 101
estandarización
en Análisis de clústeres en dos fases 226
Estimación curvilínea
almacenamiento de intervalos de predicción 205
almacenamiento de residuos 205
almacenamiento de valores pronosticados 205
análisis de la varianza 204
inclusión de constante 204
modelos 205
predicción 205
Estimaciones de Hodges-Lehman
Pruebas no paramétricas de muestras relacionadas 245
estimaciones de los parámetros
en regresión ordinal 178
estimador biponderado de Tukey
en Explorar 93
estimador en onda de Andrews
en Explorar 93
estimador-M de Huber
en Explorar 93
Estimador-M redescendente de Hampel
en Explorar 93

Estimadores M
en Explorar 93
estrés
en Escalamiento multidimensional 274
estudio de control de casos
Prueba T de muestras emparejadas 118
estudio de pares relacionados
en Prueba T de muestras emparejadas 118
eta
en Medias 103
en Tablas cruzadas 97
eta cuadrado
en Medias 103
Explorar
Características adicionales del comando 95
estadísticos 93
gráficos 94
opciones 95
transformaciones de potencia 94
valores perdidos 95

F

F múltiple de Ryan-Einot-Gabriel-Welsch
en ANOVA de un factor 123
en MLG 132
factor de inflación de la varianza
en Regresión lineal 175
factorización alfa 221
factorización de ejes principales 221
factorización de imágenes 221
fiabilidad de dos mitades
en Análisis de fiabilidad 268, 269
fiabilidad de Spearman-Brown
en Análisis de fiabilidad 269
Filtros de series temporales
Christiano Fitzgerald Filtro 202, 203
Filtro Baxter king 202, 203
Filtro Hodrick Prescott 202, 203
formato
columnas en informes 264
Frecuencias
estadísticos 88
formatos 90
gráficos 89
orden de visualización 90
supresión de tablas 90
frecuencias acumuladas
en regresión ordinal 178
frecuencias de los clústeres
en Análisis de clústeres en dos fases 227
Frecuencias de respuestas múltiples
valores perdidos 260
frecuencias esperadas
en regresión ordinal 178
frecuencias observadas
en regresión ordinal 178
funciones de densidad de probabilidad
en simulación 304
funciones de distribución acumulada
en simulación 304

G

gamma
en Tablas cruzadas 97
gamma de Goodman y Kruskal
en Tablas cruzadas 97
Generador de simulaciones 296
GLM
almacenamiento de matrices 135
gráficos de perfil 130
guardar variables 135
modelo 128
pruebas post hoc 132
suma de cuadrados 128
GLM Univariante
contrastes 129, 130
Gráfico de bosque 78
Gráfico de bosque acumulado 78
Gráfico de burbujas
Metaanálisis Regresión 80
Gráfico de embudo 78
gráfico de espacio de características
en Análisis de vecinos más próximos 213
Gráfico de Galbraith 78
Gráfico de L'Abb´e 78
gráficos
en Análisis ROC 289
en Curva ROC... 292
etiquetas de caso 204
gráficos de probabilidad 284, 286
Gráficos
Metaanálisis 78
gráficos circulares
en Frecuencias 89
gráficos de barras
en Frecuencias 89
gráficos de cargas
en Análisis factorial 222
gráficos de dispersión versus nivel
en Explorar 94
gráficos de perfil
en MLG 130
gráficos de probabilidad
P-P... 284
Q-Q 286
gráficos de probabilidad normal
en Explorar 94
en Regresión lineal 173
P-P... 284
Q-Q 286
gráficos de probabilidad P-P 284
gráficos de probabilidad Q-Q 286
gráficos de tallo y hojas
en Explorar 94
gráficos de tornado
en simulación 305
gráficos normales sin tendencia
en Explorar 94
gráficos parciales
en Regresión lineal 173
GT2 de Hochberg
en ANOVA de un factor 123
en MLG 132

H

H de Kruskal-Wallis
en Pruebas para dos muestras independientes 256
Hauck-Anderson
en proporciones de muestras independientes 115
histogramas
en Explorar 94
en Frecuencias 89
en Regresión lineal 173
historial de iteraciones
en regresión ordinal 178
homólogos
en Análisis de vecinos más próximos 214

I

importancia de variable
en Análisis de vecinos más próximos 214
importancia del predictor
modelos lineales 168
imprimir
en Bland Altman 139
información de diagnóstico de colinealidad
en Regresión lineal 175
información de diagnóstico por caso
en Regresión lineal 175
informe de estadísticos en columnas 265
intervalo de confianza
en Correlaciones bivariadas 141
Intervalos de Clopper-Pearson
Pruebas no paramétricas de una muestra 238
intervalos de confianza
almacenamiento en Regresión lineal 173
en Análisis ROC 290, 291
en ANOVA de un factor 125
en Correlaciones bivariadas 143
en Curva ROC... 293
en Explorar 93
en MLG 129
en Prueba T de muestras emparejadas 120
en Prueba t para una muestra 121
en Pruebas t para muestras independientes 118
en Regresión lineal 175
Intervalos de Jeffreys
Pruebas no paramétricas de una muestra 238
intervalos de predicción
almacenamiento en Estimación curvilínea 205
almacenamiento en Regresión lineal 173
intervalos de razón de verosimilitud
Pruebas no paramétricas de una muestra 238
Investigación 139–141
iteraciones
en Análisis de clústeres de K-medias 236
en Análisis factorial 221, 222

J

Jeffreys
en proporciones de una muestra 109

K

kappa
en Tablas cruzadas 97
kappa de Cohen
en Tablas cruzadas 97
Kappa de varios evaluadores de Fleiss 268, 269
Kappa ponderada
criterios 273
ejemplo 272
estadísticos 272, 273
imprimir 273
tabla cruzada 273
Kappa ponderada de Cohen 272, 273
KR20
en Análisis de fiabilidad 269
Kuder-Richardson 20 (KR20)
en Análisis de fiabilidad 269

L

lambda
en Tablas cruzadas 97
lambda de Goodman y Kruskal
en Tablas cruzadas 97
Lambda de Wilks
en Análisis discriminante 218
Libro de códigos
estadísticos 87
salida 85
Lineal 161
listado de casos 100
Logit
en proporciones de una muestra 109

M

mapa de cuadrantes
en Análisis de vecinos más próximos 215
matriz de configuración
en Análisis factorial 220
matriz de correlaciones
en Análisis discriminante 217
en Análisis factorial 220, 221
en regresión ordinal 178
matriz de covarianzas
en Análisis discriminante 217, 218
en MLG 135
en Regresión lineal 175
en regresión ordinal 178
matriz de transformación
en Análisis factorial 220
máxima verosimilitud
en Análisis factorial 221
máximo
comparación de columnas del informe 266
en Cubos OLAP 105
en Descriptivos 91
en Estadísticos de la razón 282
en Explorar 93
en Frecuencias 88
en Medias 103
en Resumir 101

McFadden R2
en regresión ordinal 178
McNemar
en proporciones de muestras emparejadas 112
McNemar (Continuidad corregida)
en proporciones de muestras emparejadas 112
media
de varias columnas del informe 266
en ANOVA de un factor 125
en Cubos OLAP 105
en Descriptivos 91
en el Informe de estadísticos en filas 264
en el Resúmenes de informes en columnas 266
en Estadísticos de la razón 282
en Explorar 93
en Frecuencias 88
en Medias 103
en Resumir 101
subgrupo 102, 105
media armónica
en Cubos OLAP 105
en Medias 103
en Resumir 101
media geométrica
en Cubos OLAP 105
en Medias 103
en Resumir 101
media ponderada
en Estadísticos de la razón 282
media recortada
en Explorar 93
mediana
en Cubos OLAP 105
en Estadísticos de la razón 282
en Explorar 93
en Frecuencias 88
en Medias 103
en Resumir 101
mediana agrupada
en Cubos OLAP 105
en Medias 103
en Resumir 101
Medias
estadísticos 103
opciones 103
medias de grupo 102, 105
medias de subgrupo 102, 105
medida de diferencia de configuración
en Distancias 160
medida de diferencia de tamaño
en Distancias 160
medida de disimilaridad de Lance y Williams
en Distancias 160
medida de distancia de phi cuadrado
en Distancias 160
medidas de dispersión
en Descriptivos 91
en Estadísticos de la razón 282
en Explorar 93
en Frecuencias 88
medidas de distancia
en Análisis de clústeres jerárquico 233
en Análisis de vecinos más próximos 210
en Distancias 160
medidas de distribución
en Descriptivos 91
en Frecuencias 88
medidas de similaridad
en Análisis de clústeres jerárquico 233
en Distancias 161
medidas de tendencia central
en Estadísticos de la razón 282
en Explorar 93
en Frecuencias 88
mejores subconjuntos
en modelos lineales 166
Meta-análisis binario
cuadro de diálogo Contraste 60
diálogo Análisis 58
diálogo Criterios 57
diálogo Gráfico 65
diálogo Guardar 64
diálogo Imprimir 63
diálogo Inferencia 59
diálogo Recortar y rellenar 62
diálogo Sesgo 61
pestaña Gráfico de bosque 78
pestaña Gráfico de bosque acumulado 79
Pestaña Gráfico de burbujas 80
pestaña Gráfico de embudo 81
pestaña Gráfico de Galbraith 82
pestaña Gráfico de L'Abb ´e 83
Meta-análisis continuo
cuadro de diálogo Contraste 44
diálogo Análisis 42
diálogo Criterios 41, 49
diálogo Gráfico 48
diálogo Guardar 47
diálogo Imprimir 47
diálogo Inferencia 43
diálogo Recortar y rellenar 45
diálogo Sesgo 45
pestaña Gráfico de bosque 78
pestaña Gráfico de bosque acumulado 79
Pestaña Gráfico de burbujas 80
pestaña Gráfico de embudo 81
pestaña Gráfico de Galbraith 82
pestaña Gráfico de L'Abb ´e 83
Metaanálisis
Gráficos 78
regresión 39
resultados binarios 39
resultados continuos 39
Metaanálisis Regresión
diálogo Criterios 75
diálogo Gráfico 77
diálogo Guardar 77
diálogo Imprimir 76
diálogo Inferencia 75
Gráfico de burbujas 80
mínimo
comparación de columnas del informe 266
en Cubos OLAP 105
en Descriptivos 91
en Estadísticos de la razón 282
en Explorar 93
en Frecuencias 88
en Medias 103

mínimo *(continuación)*
en Resumir 101
mínimos cuadrados generalizados
en Análisis factorial 221
mínimos cuadrados no ponderados
en Análisis factorial 221
mínimos cuadrados ponderados
en Regresión lineal 171
modelado espacial 311
modelado geoespacial 311–321
modelo compuesto
en Estimación curvilínea 205
modelo cuadrático
en Estimación curvilínea 205
modelo cúbico
en Estimación curvilínea 205
modelo de crecimiento
en Estimación curvilínea 205
modelo de curva S
en Estimación curvilínea 205
modelo de escala
en regresión ordinal 179
modelo de Guttman
en Análisis de fiabilidad 268, 269
modelo de potencia
en Estimación curvilínea 205
modelo de regresión auxiliar
en MLG 137
modelo de ubicación
en regresión ordinal 179
modelo estrictamente paralelo
en Análisis de fiabilidad 268, 269
modelo exponencial
en Estimación curvilínea 205
modelo inverso
en Estimación curvilínea 205
modelo lineal
en Estimación curvilínea 205
modelo logarítmico
en Estimación curvilínea 205
modelo logístico
en Estimación curvilínea 205
modelo paralelo
en Análisis de fiabilidad 268, 269
modelos factoriales completos
en MLG 128
modelos lineales
coeficientes 170
conjuntos 167
criterio de información 168
estadístico R cuadrado 168
importancia del predictor 168
medias estimadas 170
nivel de confianza 165
objetivos 165
opciones de Modelo 167
predicho por observado 169
preparación automática de datos 165, 168
reglas de combinación 167
réplica de resultados 167
residuos 169
resumen de generación de modelos 170
resumen del modelo 168
selección de modelos 166
modelos lineales *(continuación)*
tabla de ANOVA 169
valores atípicos 169
modelos personalizados
en MLG 128
modo
en Frecuencias 88
muestra de entrenamiento
en Análisis de vecinos más próximos 211
muestra reservada
en Análisis de vecinos más próximos 211
muestras relacionadas 255, 257
multiplicación
multiplicación entre columnas del informe 266

N

Newcombe
en proporciones de muestras emparejadas 112
en proporciones de muestras independientes 115
Newcombe (continuidad corregida)
en proporciones de muestras independientes 115
Newman-Keuls
en MLG 132
nivel de significación asintótica 269
no lineal 161
Normalización Kaiser
en Análisis factorial 222
numeración de páginas
en el informe de estadísticos en columnas 267
en informes de estadísticos en filas 264
número de casos
en Cubos OLAP 105
en Medias 103
en Resumir 101
número máximo de ramas
en Análisis de clústeres en dos fases 226

O

Omega de McDonald
en Análisis de fiabilidad 268, 269

P

pasos sucesivos hacia adelante
en modelos lineales 166
percentiles
en Explorar 93
en Frecuencias 88
en simulación 305
Percentiles 92
pestaña Gráfico de bosque
Meta-análisis binario 78
Meta-análisis continuo 78
Tamaño de efecto de meta-análisis binario 78
Tamaño de efecto de meta-análisis continuo 78
pestaña Gráfico de bosque acumulado
Meta-análisis binario 79
Meta-análisis continuo 79
Tamaño de efecto de meta-análisis binario 79
Tamaño de efecto de meta-análisis continuo 79
Pestaña Gráfico de burbujas

Pestaña Gráfico de burbujas *(continuación)*
Meta-análisis binario 80
Meta-análisis continuo 80
Tamaño de efecto de meta-análisis binario 80
Tamaño de efecto de meta-análisis continuo 80
pestaña Gráfico de embudo
Meta-análisis binario 81
Meta-análisis continuo 81
Tamaño de efecto de meta-análisis binario 81
Tamaño de efecto de meta-análisis continuo 81
pestaña Gráfico de Galbraith
Meta-análisis binario 82
Meta-análisis continuo 82
Tamaño de efecto de meta-análisis binario 82
Tamaño de efecto de meta-análisis continuo 82
pestaña Gráfico de L'Abb´e
Meta-análisis binario 83
Meta-análisis continuo 83
Tamaño de efecto de meta-análisis binario 83
Tamaño de efecto de meta-análisis continuo 83
phi
en Tablas cruzadas 97
PLUM
en regresión ordinal 176
porcentajes
en Tablas cruzadas 99
porcentajes de fila
en Tablas cruzadas 99
porcentajes de la columna
en Tablas cruzadas 99
porcentajes totales
en Tablas cruzadas 99
predicción
en Estimación curvilínea 205
preparación automática de datos
en modelos lineales 168
primero
en Cubos OLAP 105
en Medias 103
en Resumir 101
profundidad del árbol
en Análisis de clústeres en dos fases 226
Proporciones de muestras emparejadas 111
Proporciones de muestras independientes 113, 115
Proporciones de una muestra 108, 109, 112
Proximidad 143
Proximidades
en Análisis de clústeres jerárquico 232
prueba b de Tukey
en ANOVA de un factor 123
en MLG 132
prueba binomial
Pruebas no paramétricas de una muestra 238
Prueba binomial
Características adicionales del comando 250
dicotomías 249
estadísticos 250
opciones 250
valores perdidos 250
Prueba de aditividad de Tukey
en Análisis de fiabilidad 268, 269
prueba de chi-cuadrado
Pruebas no paramétricas de una muestra 238, 239
Prueba de comparación por parejas de Gabriel
Prueba de comparación por parejas de Gabriel *(continuación)*
en ANOVA de un factor 123
en MLG 132
Prueba de comparación por parejas de Games y Howell
en ANOVA de un factor 123
en MLG 132
prueba de esfericidad de Bartlett
en Análisis factorial 221
prueba de Friedman
en pruebas para varias muestras relacionadas 258
Pruebas no paramétricas de muestras relacionadas 245
prueba de homogeneidad marginal
en Pruebas para dos muestras relacionadas 255
Pruebas no paramétricas de muestras relacionadas 245
prueba de Kolmogorov-Smirnov
Pruebas no paramétricas de una muestra 238, 239
Prueba de Kolmogorov-Smirnov
Prueba de Lilliefors 239
Prueba de Kolmogorov-Smirnov para una muestra
Características adicionales del comando 253
distribución de prueba 251
estadísticos 253
opciones 253
Prueba de Lilliefors 251, 253
valores perdidos 253
prueba de la mediana
en Pruebas para dos muestras independientes 256
prueba de Levene
en ANOVA de un factor 125
en Explorar 94
prueba de Lilliefors
en Explorar 94
Prueba de Lilliefors 239, 251, 253
prueba de líneas paralelas
en regresión ordinal 178
prueba de los signos
en Pruebas para dos muestras relacionadas 255
Pruebas no paramétricas de muestras relacionadas 245
prueba de McNemar
en Pruebas para dos muestras relacionadas 255
en Tablas cruzadas 97
Pruebas no paramétricas de muestras relacionadas 245, 246
prueba de rachas
Pruebas no paramétricas de una muestra 238, 240
Prueba de rachas
Características adicionales del comando 251
estadísticos 251
opciones 251
puntos de corte 250, 251
valores perdidos 251
prueba de rangos múltiples de Duncan
en ANOVA de un factor 123
en MLG 132
prueba de reacciones extremas de Moses
en Pruebas para dos muestras independientes 254
prueba de Scheffé
en ANOVA de un factor 123
en MLG 132
prueba de Shapiro-Wilk
en Explorar 94
prueba de Wilcoxon de los rangos con signo

prueba de Wilcoxon de los rangos con signo *(continuación)*
en Pruebas para dos muestras relacionadas 255
Pruebas no paramétricas de muestras relacionadas 245
Pruebas no paramétricas de una muestra 238
prueba exacta de Fisher
en Tablas cruzadas 97
Prueba M de Box
en Análisis discriminante 217
Prueba Q de Cochran
Pruebas no paramétricas de muestras relacionadas 245, 246
Prueba t
en Prueba T de muestras emparejadas 118
en Prueba t para una muestra 120
en Pruebas t para muestras independientes 117
prueba t de Dunnett
en ANOVA de un factor 123
en MLG 132
Prueba T de muestras emparejadas
opciones 120
selección de variables emparejadas 118
valores perdidos 120
Prueba T de muestras independientes
definición de grupos 118
variables de agrupación 118
variables de cadena 118
prueba t de Sidak
en ANOVA de un factor 123
en MLG 132
prueba t de Student 117
prueba t de Waller-Duncan
en ANOVA de un factor 123
en MLG 132
prueba t dependiente
en Prueba T de muestras emparejadas 118
prueba t para dos muestras
en Pruebas t para muestras independientes 117
Prueba T para muestras independientes
intervalos de confianza 118
opciones 118
valores perdidos 118
Prueba *t* para muestras relacionadas 118
Prueba T para una muestra
Características adicionales del comando 120, 121
intervalos de confianza 121
opciones 121
valores perdidos 121
pruebas de homogeneidad de las varianzas
en ANOVA de un factor 125
pruebas de normalidad
en Explorar 94
Pruebas de Wilcoxon de una muestra 241
pruebas no paramétricas
chi-cuadrado 248
Prueba de Kolmogorov-Smirnov para una muestra 251
Prueba de rachas 250
Pruebas para dos muestras independientes 253
Pruebas para dos muestras relacionadas 255
Pruebas para varias muestras independientes 256
Pruebas para varias muestras relacionadas 257
Pruebas no paramétricas de muestras independientes
pestaña Campos 242
Pruebas no paramétricas de muestras relacionadas
Pruebas no paramétricas de muestras relacionadas *(continuación)*
campos 245
prueba de McNemar 246
Prueba Q de Cochran 246
Pruebas no paramétricas de una muestra
campos 238
prueba binomial 238
prueba de chi-cuadrado 239
prueba de Kolmogorov-Smirnov 239
prueba de rachas 240
Pruebas para dos muestras independientes
Características adicionales del comando 255
definición de grupos 255
estadísticos 255
opciones 255
tipos de pruebas 254
valores perdidos 255
variables de agrupación 255
Pruebas para dos muestras relacionadas
Características adicionales del comando 256
estadísticos 256
opciones 256
tipos de pruebas 255
valores perdidos 256
pruebas para la independencia
chi-cuadrado 97
Pruebas para varias muestras independientes
Características adicionales del comando 257
definición de rango 257
estadísticos 257
opciones 257
tipos de pruebas 257
valores perdidos 257
variables de agrupación 257
Pruebas para varias muestras relacionadas
Características adicionales del comando 258
estadísticos 258
tipos de pruebas 258
Puntuación 110
Puntuación (Continuidad corregida) 110
Puntuación Wilson
en proporciones de una muestra 109
Puntuación Wilson (continuidad corregida)
en proporciones de una muestra 109
puntuaciones factoriales 223
puntuaciones factoriales de Anderson-Rubin 223
puntuaciones factoriales de Bartlett 223
puntuaciones Z
almacenamiento como variables 90
en Descriptivos 90

Q

Q de Cochran
en pruebas para varias muestras relacionadas 258

R

R 2
Cambio en R 2 175
en Medias 103
en Regresión lineal 175
R 2 corregida

R 2 corregida *(continuación)*
en Regresión lineal 175
R múltiple
en Regresión lineal 175
R-cuadrado
en modelos lineales 168
R-cuadrado corregida
en modelos lineales 166
R-E-G-W F
en ANOVA de un factor 123
en MLG 132
R-E-G-W Q
en ANOVA de un factor 123
en MLG 132
R2 de Cox y Snell
en regresión ordinal 178
R2 de Nagelkerke
en regresión ordinal 178
Rachas de Wald-Wolfowitz
en Pruebas para dos muestras independientes 254
rango
en Cubos OLAP 105
en Descriptivos 91
en Estadísticos de la razón 282
en Frecuencias 88
en Medias 103
en Resumir 101
rango múltiple de Ryan-Einot-Gabriel-Welsch
en ANOVA de un factor 123
en MLG 132
razón entre covarianzas
en Regresión lineal 173
recuento esperado
en Tablas cruzadas 99
recuento observado
en Tablas cruzadas 99
reglas de combinación
en modelos lineales 167
regresión
gráficos 173
Metaanálisis 39
Regresión lineal 171
regresión múltiple 171
Regresión lineal
bloques 171
Características adicionales del comando 176
estadísticos 175
exportación de información del modelo 173
gráficos 173
guardar variables nuevas 173
métodos de selección de variables 172, 175
ponderaciones 171
residuos 173
valores perdidos 175
variable de selección 173
Regresión lineal de Elastic Net
opciones 181
partición 180
variables 180
Regresión lineal de Lasso
opciones 184
partición 184
variables 184
Regresión lineal Ridge
Regresión lineal Ridge *(continuación)*
opciones 188
partición 187
variables 187
regresión múltiple
en Regresión lineal 171
Regresión ordinal
Características adicionales del comando 180
enlace 177
estadísticos 176
modelo de escala 179
modelo de ubicación 179
opciones 177
Regresión por mínimos cuadrados parciales
exportar variables 207
modelo 207
relación espacial 143
residuos
almacenamiento en Estimación curvilínea 205
almacenamiento en Regresión lineal 173
en Tablas cruzadas 99
residuos de Pearson
en regresión ordinal 178
residuos eliminados
en MLG 135
en Regresión lineal 173
residuos estandarizados
en MLG 135
en Regresión lineal 173
residuos estudentizados
en Regresión lineal 173
residuos no estandarizados
en MLG 135
Respuesta múltiple
Características adicionales del comando 262
resultados binarios
datos de tamaño de efecto calculados previamente 39
datos de tamaño de efecto en bruto 39
resultados continuos
datos de tamaño de efecto calculados previamente 39
datos de tamaño de efecto en bruto 39
Metaanálisis 39
resumen de error
en Análisis de vecinos más próximos 216
Resúmenes de informes en columnas
Características adicionales del comando 267
columnas totales 266
control de página 267
diseño de página 265
formato de columnas 264
numeración de páginas 267
subtotales 267
total final 267
valores perdidos 267
Resúmenes de informes en filas
Características adicionales del comando 267
columnas de datos 263
control de página 264
diseño de página 265
espaciado de salto 264
formato de columnas 264
numeración de páginas 264
ordenación de secuencias 263
pies 265

Resúmenes de informes en filas *(continuación)*
salto de columna 263
títulos 265
valores perdidos 264
variables en los títulos 265
Resumir
estadísticos 101
opciones 101
rho
en Correlaciones bivariadas 141
en Tablas cruzadas 97
riesgo
en Tablas cruzadas 97
riesgo relativo
en Tablas cruzadas 97
rotación equamax
en Análisis factorial 222
rotación oblimin directa
en Análisis factorial 222
rotación quartimax
en Análisis factorial 222
rotación varimax
en Análisis factorial 222

S

S-stress
en Escalamiento multidimensional 274
selección de características
en Análisis de vecinos más próximos 215
selección de características y k
en Análisis de vecinos más próximos 215
selección de k
en Análisis de vecinos más próximos 215
selección hacia delante
en Análisis de vecinos más próximos 210
en Regresión lineal 172
selección por pasos
en Regresión lineal 172
simulación
ajuste de distribución 299
análisis de sensibilidad 302
análisis hipotético 302
correlaciones entre entradas 303
creación de nuevas entradas 298
creación de un plan de simulación 294, 295
criterios de parada 303
diagramas de cajas 305
diagramas de dispersión 305
editor de ecuaciones 297
ejecución de un plan de simulación 295, 307
especificación de modelo 296
formatos de visualización de destinos y entradas 305
función de densidad de probabilidad 304
función de distribución acumulada 304
Generador de simulaciones 296
gráficos de tornado 305
gráficos interactivos 309
guardar datos simulados 306
guardar plan de simulación 306
modelos admitidos 296
muestreos de cola 303
opciones del gráfico 310
percentiles de distribuciones de destino 305
simulación *(continuación)*
personalización del ajuste de distribución 301
reajuste de distribuciones a nuevos datos 307
resultados de ajuste de distribución 301
salida 304, 305
Simulación de Monte Carlo 293
Student-Newman-Keuls
en ANOVA de un factor 123
en MLG 132
subtotales
en el informe de estadísticos en columnas 267
suma
en Cubos OLAP 105
en Descriptivos 91
en Frecuencias 88
en Medias 103
en Resumir 101
suma de cuadrados
en MLG 128

T

T 2 de Hotelling
en Análisis de fiabilidad 268, 269
T2 de Tamhane
en ANOVA de un factor 123
en MLG 132
T3 de Dunnett
en ANOVA de un factor 123
en MLG 132
tabla cruzada
en Tablas cruzadas 95
respuesta múltiple 261
tabla de clasificación
en Análisis de vecinos más próximos 215
Tablas cruzadas
capas 96
estadísticos 97
formatos 100
gráficos de barras agrupadas 97
presentación de casillas 99
supresión de tablas 95
variables de control 96
Tablas cruzadas de respuestas múltiples
definición de rangos de valores 262
emparejamiento de las variables entre los conjuntos de respuestas 262
porcentajes basados en casos 262
porcentajes basados en respuestas 262
porcentajes de casillas 262
valores perdidos 262
tablas de contingencia 95
tablas de frecuencias
en Explorar 93
en Frecuencias 87
tamaño de efecto
en prueba T de muestras emparejadas 118
en Pruebas t para muestras independientes 117
Tamaño de efecto de meta-análisis binario
cuadro de diálogo Contraste 69
diálogo Análisis 67
diálogo Criterios 66
diálogo Gráfico 73
diálogo Guardar 73

Tamaño de efecto de meta-análisis binario *(continuación)*
diálogo Imprimir 72
diálogo Inferencia 68
diálogo Recortar y rellenar 70
diálogo Sesgo 70
pestaña Gráfico de bosque 78
pestaña Gráfico de bosque acumulado 79
Pestaña Gráfico de burbujas 80
pestaña Gráfico de embudo 81
pestaña Gráfico de Galbraith 82
pestaña Gráfico de L'Abb´e 83
Tamaño de efecto de meta-análisis continuo
cuadro de diálogo Contraste 52
diálogo Análisis 50
diálogo Gráfico 56
diálogo Guardar 55
diálogo Imprimir 54
diálogo Inferencia 51
diálogo Recortar y rellenar 53
diálogo Sesgo 52
pestaña Gráfico de bosque 78
pestaña Gráfico de bosque acumulado 79
Pestaña Gráfico de burbujas 80
pestaña Gráfico de embudo 81
pestaña Gráfico de Galbraith 82
pestaña Gráfico de L'Abb´e 83
tau de Goodman y Kruskal
en Tablas cruzadas 97
tau de Kruskal
en Tablas cruzadas 97
tau-b
en Tablas cruzadas 97
Tau-b de Kendall
en Correlaciones bivariadas 141
en Tablas cruzadas 97
tau-c
en Tablas cruzadas 97
Tau-c de Kendall
en Tablas cruzadas 97
términos de interacción 128, 179, 180
títulos
en Cubos OLAP 107
tolerancia
en Regresión lineal 175
Tolerancia
Regresión lineal 175
totales finales
en el informe de estadísticos en columnas 267
tratamiento del ruido
en Análisis de clústeres en dos fases 226
trazos
en Bland Altman 139
TSF 202, 203

U

U de Mann-Whitney
en Pruebas para dos muestras independientes 254
umbral inicial
en Análisis de clústeres en dos fases 226

V

V
en Tablas cruzadas 97
V de Cramér
en Tablas cruzadas 97
V de Rao
en Análisis discriminante 218
valores atípicos
en Análisis de clústeres en dos fases 226
en Explorar 93
en Regresión lineal 173
valores de influencia
en MLG 135
en Regresión lineal 173
valores estandarizados
en Descriptivos 90
valores extremos
en Explorar 93
valores perdidos
en Análisis de vecinos más próximos 212
en Análisis factorial 223
en Análisis ROC 290, 291
en ANOVA de un factor 125
en Correlaciones bivariadas 142
en Correlaciones parciales 159
en Curva ROC... 293
en el informe de estadísticos en columnas 267
en el Informe de estadísticos en filas 264
en Explorar 95
en la prueba de chi-cuadrado 249
en las frecuencias de respuestas múltiples 260
en las tablas cruzadas de respuestas múltiples 262
en proporciones de muestras emparejadas 113
en proporciones de muestras independientes 116
en proporciones de una muestra 110
en Prueba binomial 250
en Prueba de Kolmogorov-Smirnov de una muestra 253
en Prueba de rachas 251
en Prueba T de muestras emparejadas 120
en Prueba t para una muestra 121
en Pruebas para dos muestras independientes 255
en Pruebas para dos muestras relacionadas 256
en Pruebas para varias muestras independientes 257
en Pruebas t para muestras independientes 118
en Regresión lineal 175
Valores perdidos
en percentiles 92
valores pronosticados
almacenamiento en Estimación curvilínea 205
almacenamiento en Regresión lineal 173
valores pronosticados ponderados
en MLG 135
variable de selección
en Regresión lineal 173
variables de control
en Tablas cruzadas 96
varianza
en Cubos OLAP 105
en Descriptivos 91
en el Informe de estadísticos en filas 264
en el Resúmenes de informes en columnas 266
en Explorar 93

varianza *(continuación)*
en Frecuencias 88
en Medias 103
en Resumir 101
visor de clústeres
clasificación de la visualización de características 229
clasificación de la visualización de clústeres 230
clasificar características 229
clasificar clústeres 230
clasificar contenido de casillas 230
comparación de clústeres 231
distribución de casillas 231
filtrado de registros 232
importancia del predictor 230
información sobre los modelos de clúster 227
resumen del modelo 228
tamaño de los clústeres 230
transponer clústeres y características 229
utilización 231
visión general 228
vista básica 230
vista comparación de clústeres 231
vista de centros de clústeres 229
vista de clústeres 229
vista de resumen 228
vista de tamaños de clústeres 230
vista distribución de casillas 231
vista importancia del predictor de clústeres 230
visualización de contenido de casillas 230
voltear clústeres y características 229
vista de modelo
en Análisis de vecinos más próximos 213
visualización
modelos de clústeres 228

W

W de Kendall
en pruebas para varias muestras relacionadas 258
Wald
en proporciones de muestras emparejadas 112
en proporciones de muestras independientes 115
en proporciones de una muestra 109
Wald (continuidad corregida)
en proporciones de muestras emparejadas 112
en proporciones de muestras independientes 115
en proporciones de una muestra 109
Wald (Continuidad corregida)
en proporciones de muestras emparejadas 112
en proporciones de muestras independientes 115
Wald H0
en proporciones de muestras independientes 115
Wald H0 (Continuidad corregida)
en proporciones de muestras independientes 115

X

X^2 139–141

Z

Z de Kolmogorov-Smirnov
en Prueba de Kolmogorov-Smirnov de una muestra 251
en Pruebas para dos muestras independientes 254