IBM SPSS Statistics Base 32

Nota

Prima di utilizzare queste informazioni e il prodotto che supportano, leggere le informazioni in “Informazioni particolari” a pagina 317.

Informazioni sul prodotto

Questa edizione si applica alla versione 32, release 0, modifica 0 di IBM® SPSS Statistica e a tutte le release e modifiche successive se non diversamente indicato in nuove edizioni.

© Copyright International Business Machines Corporation .

Indice

Capitolo 1. Funzioni principali....1
Analisi di potenza....1
Medie....2
Proporzioni....16
Correlazioni....26
Regressione....35
Analisi di alimentazione: Precisione....38
Analisi di alimentazione: Valori Grigi....38
Meta - analisi....39
Meta - Analisi Continuo....39
Meta - Analisi Continua Effetto Dimensione....47
Binario meta - analisi....55
Meta - Analisi Effetto Binario Dimensione....64
Regressione meta - analisi....73
Opzioni di tracciamento meta - analisi....77
Informazioni sui dati....83
Scheda Output della finestra Informazioni sui dati....84
Scheda Informazioni sui dati - Statistiche....86
Frequenze....86
Frequenze: Statistiche....87
Frequenze: Grafici....88
Frequenze: Formato....89
Descrittive....89
Descrittive: Opzioni....89
Funzioni aggiuntive del comando DESCRIPTIVES....90
Percentili....90
Valori mancanti percentili....91
Esplora....91
Esplora: Statistica....92
Esplora: Grafici....93
Esplora: Opzioni....94
Funzioni aggiuntive del comando EXAMINE....94
Tabelle di contingenza....94
Strati delle tabelle di contingenza....95
Grafici a barre raggruppate di tabelle di contingenza....95
Tabelle di contingenza che visualizzano le variabili di strato negli strati della tabella....96
Statistiche delle tabelle di contingenza....96
Visualizzazione celle delle tabelle di contingenza....97
Formato della tabella di tavole di contingenza....99
Riepiloga....99
Riassumi: Opzioni....99
Riassumi: Statistiche....100
Medie....101
Medie: Opzioni....102
Cubi OLAP....103
Cubi OLAP: Statistiche....104
Cubi OLAP: Differenze....106
Cubi OLAP: Titolo....106
Proporzioni....106
Introduzione delle proporzioni....106
Proporzioni campione singolo....106

Proporzioni campioni accoppiati..... 109
Proporzioni campioni indipendenti.....112
Test T.....115
Test T..... 115
Test T a campioni indipendenti.....115
Test T a campioni accoppiati..... 117
Test T a campione singolo.....118
Funzioni aggiuntive di comando t TEST.....119
ANOVA a una via..... 120
ANOVA a una via: Contrasti.....121
ANOVA a una via: Test Post Hoc..... 121
ANOVA a una via: Opzioni..... 123
Funzioni aggiuntive del comando ONEWAY..... 124
Analisi GLM univariato..... 124
GLM – Univariato: Modello.....126
GLM – Univariato: Contrasti..... 127
GLM – Univariato: Grafici di profilo.....128
GLM: Confronti post hoc.....130
GLM – Univariato: Salva..... 133
Medie marginali stimate GLM..... 134
GLM – Univariato: Opzioni.....134
Funzioni aggiuntive del comando UNIANOVA.....135
Bland-Altman.....136
Bland Altman: criteri..... 136
Bland Altman: Stampa..... 137
Bland Altman: trama..... 137
Test del chi-quadrato..... 137
Test del chi-quadro: Esatta.....138
Test del chi-quadro: Celle..... 138
Test del chi-quadro: Formato..... 138
Correlazioni di bivariate.....138
Correlazioni bivariate: Opzioni.....140
Intervallo Di Confidenza Correlazioni Bivariate.....140
Funzioni aggiuntive dei comandi CORRELATIONS e NONPAR CORR..... 141
Associazione di prossimità..... 141
Mappatura di prossimità: Dati.....143
Mappatura di prossimità: Pesi..... 143
Mappatura di prossimità: Iniziale..... 144
Mappatura di prossimità: Modello.....145
Mappatura di prossimità: Restrizioni.....147
Mappatura di prossimità: Supplemento.....148
Mappatura di prossimità: Criteri..... 148
Mappatura di prossimità: Uscita.....150
Mappatura di prossimità: Salva..... 155
Correlazioni parziali..... 156
Correlazioni parziali: Opzioni..... 157
Funzioni aggiuntive del comando PARTIAL CORR.....157
Distanze..... 157
Distanze: Misure di dissimilarità.....158
Distanze: Misure di similarità.....158
Funzioni aggiuntive del comando PROXIMITIES..... 159
Correlazione della distanza..... 159
Correlazione della distanza: Criteri.....160
Correlazione della distanza: Stampa..... 161
Correlazione della distanza: grafico.....161
Modelli lineari..... 162
Come ottenere un modello lineare..... 162
Obiettivi 162

Di base ... 163
Selezione modello ... 163
Insiemi ... 164
Avanzate ... 165
Opzioni modello ... 165
Riepilogo del modello ... 165
Preparazione automatica dati ... 165
Importanza predittore ... 166
Previsioni e osservazioni ... 166
Residui ... 166
Valori anomali ... 166
Effetti ... 166
Coefficienti ... 167
Medie stimate ... 168
Riepilogo creazione modello ... 168
Regressione lineare ... 168
Metodi di selezione della variabile di regressione lineare ... 169
Regressione lineare: Imposta regola ... 170
Regressione lineare: grafici ... 170
Regressione lineare: salvataggio di nuove variabili ... 170
Regressione lineare: Statistiche ... 172
Regressione lineare: Opzioni ... 173
Funzioni aggiuntive del comando REGRESSION ... 173
Regressione ordinale ... 173
Regressione ordinale: Opzioni ... 174
Regressione ordinale: Output ... 175
Regressione ordinale: Ubicazione ... 176
Regressione ordinale: Scala ... 176
Funzioni aggiuntive del comando PLUM ... 177
Regressione Elastic Net lineare ... 177
Linear Elastic Net Regression: Opzioni ... 178
Regressione Lasso lineare ... 181
Regressione Lineare Laica: Opzioni ... 181
Regressione Lineare Ridge ... 184
Regressione Lineare Ridge: Opzioni ... 184
Analisi di mediazione ... 187
Tipo di modello ... 192
Variabili dipendenti ... 193
Predittore ... 194
Criteri ... 195
Stampa ... 196
Grafici ... 197
Filtri per serie temporali ... 199
Filtri per serie temporali: Opzioni ... 199
Stima di curve ... 200
Stima di curve: Modelli ... 201
Stima di curve: Salva ... 202
Regressione dei minimi quadrati parziali ... 202
Modello ... 203
Opzioni ... 204
Analisi del vicino più vicino ... 204
Vicini ... 206
Funzioni ... 207
Partizioni ... 207
Salva ... 208
Output ... 208
Opzioni ... 209
Vista Modello ... 209

Analisi discriminante ... 212
Analisi discriminante: Definisci intervallo ... 213
Analisi discriminante: Seleziona casi ... 213
Analisi discriminante: Statistiche ... 213
Analisi discriminante: Metodo stepwise ... 214
Analisi discriminante: Classificazione ... 215
Analisi discriminante: Salva ... 215
Funzioni aggiuntive del comando DISCRIMINANT ... 216
Analisi fattoriale ... 216
Analisi fattoriale: Seleziona casi ... 217
Analisi fattoriale: Descrittive ... 217
Analisi fattoriale: Estrazione ... 217
Analisi fattoriale: Rotazione ... 218
Analisi fattoriale: Punteggi fattoriali ... 219
Analisi fattoriale: Opzioni ... 219
Funzioni aggiuntive del comando FACTOR ... 219
Scelta di una procedura per il raggruppamento ... 220
Analisi cluster TwoStep ... 220
Opzioni di Analisi cluster TwoStep ... 222
Output di Analisi cluster TwoStep ... 223
Il Visualizzatore cluster ... 223
Analisi cluster gerarchica ... 228
Analisi cluster gerarchica: Metodo ... 229
Analisi cluster gerarchica: Statistiche ... 230
Analisi cluster gerarchica: Grafici ... 230
Analisi cluster gerarchica: Salva nuove variabili ... 230
Funzioni aggiuntive della sintassi del comando CLUSTER ... 230
Analisi del cluster delle k-medie ... 230
Efficienza dell'analisi del cluster delle k-medie ... 231
Analisi del cluster delle k-medie: Itera ... 231
Analisi del cluster delle k-medie: Salva ... 232
Analisi del cluster delle k-medie: Opzioni ... 232
Funzioni aggiuntive del comando QUICK CLUSTER ... 232
Test non parametrici ... 233
Test non parametrici a campione singolo ... 233
Test di Wilcoxon a campione singolo ... 237
Test non parametrici di campioni indipendenti ... 237
Test non parametrici di campioni correlati ... 240
Funzioni aggiuntive del comando NPTESTS ... 242
Finestre legacy ... 243
Analisi a risposta multipla ... 254
Analisi a risposta multipla ... 254
Risposte multiple: Definisci insiemi ... 254
Risposte multiple: Frequenze ... 255
Risposte multiple: Tabelle di contingenza ... 256
Risultati di report ... 258
Risultati di report ... 258
Report : Riepiloghi per righe ... 258
Report: Riepiloghi per colonne ... 260
Funzioni aggiuntive del comando REPORT ... 263
Analisi di affidabilità ... 263
Analisi di affidabilità: Statistiche ... 264
Funzioni aggiuntive del comando RELIABILITY ... 267
Kappa pesata ... 267
Kappa pesata: criteri ... 268
Kappa pesata: stampa ... 268
Scaling multidimensionale ... 269
Scaling multidimensionale: Forma dei dati ... 270

Scaling multidimensionale: Crea misure dai dati....270
Scaling multidimensionale: Modello....271
Scaling multidimensionale: Opzioni....271
Funzioni aggiuntive del comando ALSCAL....271
Confronto tra ALSCAL e PROXMAP....272
Statistiche dei rapporti....277
Statistiche dei rapporti: Statistiche....277
Grafico P-P....279
Grafico Q-Q....281
Analisi ROC....284
Analisi ROC: Opzioni....285
Analisi ROC: Visualizzazione....286
Analisi ROC: Definire i gruppi (stringa)....286
Analisi ROC: Definire i gruppi (numerici)....287
Curva ROC....287
Curva ROC: Opzioni....288
Simulazione....288
Progettazione di una simulazione in base a un file del modello....289
Progettazione di una simulazione in base a equazioni personalizzate....289
Progettazione di una simulazione senza un modello predittivo....290
Eseguire una simulazione da un piano di simulazione....290
Builder di simulazioni....291
Finestra di dialogo Esegui simulazione....302
Utilizzo dell'output del grafico dalla simulazione....304
Modellazione geospaziale....305
Selezione di mappe....306
Origini dati....308
Regole di associazione geospaziale....309
Previsione temporale spaziale....312
Fine....316

Informazioni particolari....317
Marchi....318

Indice analitico....321

Capitolo 1. Funzioni principali

Le seguenti funzioni principali sono incluse in IBM SPSS Statistica Base Edition.

Analisi di potenza

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

IBM SPSS Statistica fornisce le seguenti procedure di Power Analysis:

Test T a campione singolo
In un'analisi a campione, i dati osservati vengono raccolti come un singolo campione casuale. Si suppone che i dati di esempio seguano in modo indipendente e identicamente una normale distribuzione con media e varianza fissa, e richiama inferenza statistica sul parametro medio.

Test T-Test accoppiato
In analisi a campione, i dati osservati contengono due campioni accoppiati e correlati e ogni caso ha due misurazioni. Si suppone che i dati in ogni campione seguano in modo indipendente e identicamente una normale distribuzione con una media e una varianza fissa, e richiama inferenza statistica sulla differenza dei due mezzi.

Test T-Test indipendente
Nell'analisi a campione indipendente, i dati osservati contengono due campioni indipendenti. Si suppone che i dati in ogni campione seguano in modo indipendente e identicamente una normale distribuzione con una media e una varianza fissa, e richiama inferenza statistica sulla differenza dei due mezzi.

ANOVA a una via
L'analisi della varianza (ANOVA) è un metodo statistico di stima dei mezzi di diverse popolazioni che spesso si presume siano normalmente distribuiti. La One-way ANOVA, un tipo comune di ANOVA, è un'estensione del test a due sample *t*.

Esempio. La potenza di un test di ipotesi per rilevare un'ipotesi alternativa corretta è la probabilità che il test rigetta l'ipotesi del test. Dato che la probabilità di un errore di tipo II è la probabilità di accettare l'ipotesi nulla quando l'ipotesi alternativa è vera, la potenza può essere espressa come (1-probability di un errore di tipo II), ovvero la probabilità di rifiutare l'ipotesi nulla quando l'ipotesi alternativa è vera.

Statistiche e grafici. Prova unilaterale, test a due lati, livello di significatività, tasso di errore di tipo I, ipotesi di prova, deviazione standard della popolazione, media di popolazione in fase di test, valore ipotizzato, potenza bidimensionale per dimensione del campione, potenza tridimensionale per dimensione del campione, potenza tridimensionale per dimensione, gradi di rotazione, coppie di gruppi, coefficiente di correlazione del prodotto - momento, differenza media, intervallo grafico della dimensione dell'effetto, deviazione standard della popolazione, contrasti e pairwise, coefficienti di contrasto, test di contrasto, test di contrasto, BONFERRONI, SIDAK, LSD, potenza per dimensione totale del campione, potenza bidimensionale per deviazione standard pooled, potenza tridimensionale per campione totale, potenza tridimensionale per dimensione totale del campione, deviazione standard pooled, distribuzione t - distribuzione di Student, non centrale t - distribuzione,

Considerazioni sui dati dell'analisi di

Dati
In un'analisi a campione, i dati osservati vengono raccolti come un singolo campione casuale.

In analisi a campione, i dati osservati contengono due campioni accoppiati e correlati e ogni caso ha due misurazioni.

Nell'analisi a campione indipendente, i dati osservati contengono due campioni indipendenti.

L'analisi della varianza (ANOVA) è un metodo statistico di stima dei mezzi di diverse popolazioni che spesso si presume siano normalmente distribuiti.

Ipotesi

In un'analisi a campione, si suppone che i dati di esempio seguano in modo indipendente e identicamente una normale distribuzione con una media e una varianza fissa, e richiama inferenza statistica sul parametro medio.

Nell'analisi dei campioni in coppia si ipotizza che i dati in ogni campione seguano in modo indipendente e identicamente una normale distribuzione con una media e una varianza fissa, e richiama inferenza statistica sulla differenza dei due mezzi.

Nell'analisi a campione indipendente si ipotizza che i dati in ogni campione seguano in modo indipendente e identicamente una normale distribuzione con una media e una varianza fissa, e richiama inferenza statistica sulla differenza dei due mezzi.

In senso unico ANOVA, il metodo statistico di stima dei mezzi di diverse popolazioni è spesso ipotizzabile normalmente distribuito.

Ottenimento di una Power Analysis

1. Dai menu, scegliere:

   Analizzare > **Analisi di alimentazione** > **Confronta significa** > **One - Sample T-Test**, oppure **Pandate - sample T-Test**, oppure **Test T-Test indipendente**o **One-way ANOVA**
2. Definire le ipotesi di test richieste.
3. Fare clic su **OK**.

Medie

Le seguenti funzioni statistiche sono incluse in IBM SPSS Statistica Base Edition.

Analisi di potenza del test T a campione singolo

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

In un'analisi a campione, i dati osservati vengono raccolti come un singolo campione casuale. Si suppone che i dati di esempio seguano in modo indipendente e identicamente una normale distribuzione con media e varianza fissa, e richiama inferenza statistica sul parametro medio.

1. Dai menu, scegliere:

   Analizzare > **Analisi Alimentazione** > **Mezzi** > **One - sample T Test**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

Per ulteriori informazioni, consultare “Analisi di alimentazione: Valori Grigi” a pagina 38.

4. Quando **Power** viene selezionato come metodo di test supposizione **Stima** , inserire un'appropriata **Dimensione di esempio** per il valore di stima dell'alimentazione. Il valore deve essere un numero intero maggiore di 1.
5. Facoltativamente, selezionare un'opzione dal campo **Specificare** .

   Valori ipotizzati
   L'impostazione predefinita fornisce le impostazioni **Medie Popolari** e **Valore Null** .

   Media della popolazione
   Inserire un valore che specifica la popolazione media sotto prova. Il valore deve essere un singolo numerico.

   Valore null
   Facoltativamente, inserire un valore che specifica il valore di ipotesi nullo da verificare. Il valore deve essere un singolo numerico.

   Dimensione effetto
   Il f di Cohen viene utilizzato per stimare la dimensione dell'effetto come input per la stima della potenza o della dimensione del campione. La dimensione dell'effetto definito **Valore** viene passata al passo intermedio nella procedura e calcola la potenza desiderata o la dimensione del campione.
6. Inserire un valore **Deviazione Standard di Popolazione** . Il valore deve essere un singolo numerico maggiore di 0.
7. Selezionare se il test è uno o due lati.

   Analisi non direzionale (bidirezionale)
   Quando viene selezionato, viene utilizzato un test a due lati. Questa è l'impostazione predefinita.

   Analisi direzionale (unilaterale)
   Quando selezionato, la potenza viene calcolata per un test unilaterale.
8. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.
9. È possibile opzionalmente fare clic su **Plot** per specificare le impostazioni “Analisi alimentazione di One - Esempio T Test: Plot” a pagina 3 (output grafico, impostazioni della trama bidimensionale e impostazioni della trama tridimensionale).

   Nota: Plot è disponibile solo quando **Power** viene selezionato come presupposto di prova.
10. Opzionalmente, clicca su **Precisione** per stimare la dimensione del campione in base agli intervalli di confidenza specificando i valori dell'intervallo di confidenza di mezza larghezza. Per ulteriori informazioni, consultare “Analisi di alimentazione: Precisione” a pagina 38.

Nota: Precision è disponibile solo quando **Dimensione di esempio** viene selezionata come metodo di test di test **Stima** , **Valori Hypothesize** è selezionato dall'elenco **Specificare** e **L'analisi non direzionale (a due lati)** viene selezionata come **Test Direction**.

Analisi alimentazione di One - Esempio T Test: Plot

La finestra di dialogo **Plot** fornisce opzioni per il controllo dei complotti che sono in uscita per illustrare la potenza a due e tridimensionali mediante grafici di dimensioni di esempio ed effetto. La finestra di dialogo controlla anche i gradi di rotazione verticali e orizzontali per i grafici tridimensionali.

Grafico a due dimensioni

Confronto tra stima della potenza e dimensione del campione
Quando è abilitata, questa impostazione opzionale fornisce le opzioni per il controllo della potenza bidimensionale per dimensione del campione. L'impostazione è disabilitata per impostazione predefinita.

Intervallo dimensione campione
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione del campione.

Limite inferiore
Controlla il limite inferiore per la potenza bidimensionale per dimensione del campione. Il valore deve essere maggiore di 1 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la potenza bidimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Confronto tra stima della potenza e dimensione degli effetti
Per impostazione predefinita, questa impostazione opzionale è disabilitata. Quando è abilitata, il grafico visualizza nell'output. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , la gamma di tracciamento predefinita della dimensione di effetto utilizzata.

Intervallo della dimensione degli effetti
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori.

Limite inferiore
Controlla il limite inferiore per la potenza bidimensionale per effetto grafico ad effetto. Il valore deve essere maggiore o uguale a -5.0 e non può essere maggiore del valore **Limite superiore** .

Limite superiore
Controlla il limite superiore per la potenza bidimensionale per effetto grafico ad effetto. Il valore deve essere maggiore del valore **Limit. inferiore** e non può essere maggiore di 5.0.

Grafico a tre dimensioni

Confronto tra stima della potenza e
Fornisce opzioni per il controllo della potenza tridimensionale per dimensione del campione (x - asse) e dimensioni dell'effetto (asse y), le impostazioni di rotazione verticale / orizzontale e la gamma di tracciamento specificato dall'utente. Per impostazione predefinita, questa impostazione è disabilitata.

Dimensione degli effetti sull'asse x e dimensione del campione sull'asse y
L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (x - asse) e dimensioni dell'effetto (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Dimensione degli effetti sull'asse y e dimensione del campione sull'asse x
L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (asse y) e dimensione dell'effetto (asse x). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Intervallo dimensione campione
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione del campione.

Limite inferiore
Controlla il limite inferiore per la potenza tridimensionale per dimensione del campione. Il valore deve essere maggiore di 1 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la potenza tridimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Intervallo della dimensione degli effetti
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione dell'effetto.

Limite inferiore
Controlla il limite inferiore per la potenza tridimensionale per effetto grafico ad effetto. Il valore deve essere maggiore o uguale a -5.0 e non può essere maggiore del valore **Limite superiore** .

Limite superiore
Controlla il limite superiore per la potenza tridimensionale per effetto grafico ad effetto. Il valore deve essere maggiore del valore **Limit. inferiore** e non può essere maggiore di 5.0.

Analisi di potenza del test T a campioni accoppiati

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

Nell'analisi dei campioni in coppia, i dati osservati contengono due campioni accoppiati e correlati e ogni caso ha due misurazioni. Si suppone che i dati in ogni campione seguano in modo indipendente e identicamente una normale distribuzione con una media e una varianza fissa, e richiama inferenza statistica sulla differenza dei due mezzi.

1. Dai menu, scegliere:

   Analizzare > **Analisi di alimentazione** > **medie** > **Pandate - campioni t test**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

   Per ulteriori informazioni, consultare “Analisi di alimentazione: Valori Grigi” a pagina 38.
4. Quando **Power** viene selezionato come metodo di test supposizione **Stima** , inserire il valore appropriato **Dimensione di esempio** . Il valore deve essere un intero positivo maggiore di o uguale a 2.
5. Facoltativamente, selezionare un'opzione dal campo **Specificare** .

   Valori ipotizzati
   L'impostazione predefinita fornisce le impostazioni **Mean** e **Standard Deviazione** .

   Differenza media di popolazione
   Quando si richiede una singola media di popolazione, inserire un valore **Differenza media di popolazione** . Quando viene specificato un solo valore, denota la differenza media μ_d.

Nota: Il valore non può essere 0 quando si seleziona **Dimensione Esempio** .

Popolazione media per gruppo 1 e gruppo 2
Quando sono necessari più mezzi di popolazione per le coppie di gruppo specificate, inserire valori per **Popolazione Media per il gruppo 1** e **Popolazione media per il gruppo 2**. Quando vengono specificati più valori, essi denotano la differenza media di popolazione μ_1 e μ_2.

Nota: I due valori non possono essere gli stessi quando viene selezionata **Dimensione di esempio** .

Deviazione standard della popolazione per differenza media
Quando si specifica una singola media di popolazione, inserire la deviazione standard della popolazione per il valore di differenza medio. Quando viene specificato un solo valore, denota la deviazione standard della popolazione della differenza di gruppo σ_d. Il valore deve essere un singolo numerico maggiore di 0.

Deviazione standard della popolazione per il gruppo 1 e il gruppo 2
Quando si specificano più mezzi di popolazione, inserire la deviazione standard della popolazione per i valori di gruppo 1 e di gruppo 2. Quando vengono specificati più valori, essi denotano la deviazione standard della popolazione della differenza di gruppo σ_1 e σ_2. I valori devono essere un numero singolo superiore a 0.

Coefficiente di correlazione del prodotto - momento di Pearson
Opzionalmente, inserire un valore che specifica il coefficiente di correlazione del prodotto di Pearson ρ. Il valore deve essere un singolo valore numerico compreso tra -1 e 1. Il valore non può essere 0.

Nota: Quando viene specificato un singolo valore di **Popolazione Deviazione Standard per la differenza media** , questa impostazione viene ignorata. In caso contrario, i valori per **La deviazione standard della popolazione per il gruppo 1 e il gruppo 2** sono utilizzati per calcolare σ_d.

Dimensione effetto
Stima la dimensione dell'effetto come input per la stima della potenza o della dimensione del campione. La dimensione dell'effetto definito **Valore** viene passata al passo intermedio nella procedura e calcola la potenza desiderata o la dimensione del campione.

6. Selezionare se il test è uno o due lati.

   Analisi non direzionale (bidirezionale)
   Quando viene selezionato, viene utilizzato un test a due lati. Questa è l'impostazione predefinita.

   Analisi direzionale (unilaterale)
   Quando selezionato, la potenza viene calcolata per un test unilaterale.

7. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.

8. Opzionalmente, clicca su **Plot** per specificare le impostazioni “Analisi alimentazione di Pandate - Samples T Test: Plot” a pagina 6 (output grafico, impostazioni della trama bidimensionale, impostazioni della trama tridimensionale e suggerimenti).

   Nota: Plot è disponibile solo quando **Power** viene selezionato come presupposto di prova.

9. Opzionalmente, clicca su **Precisione** per stimare la dimensione del campione in base agli intervalli di confidenza specificando i valori dell'intervallo di confidenza di mezza larghezza. Per ulteriori informazioni, consultare “Analisi di alimentazione: Precisione” a pagina 38.

Nota: Precision è disponibile solo quando **Dimensione di esempio** viene selezionata come metodo di test supposizione **Stima** e **Valori Hypothesize** viene selezionato dall'elenco **Specificare** .

Analisi alimentazione di Pandate - Samples T Test: Plot

La finestra di dialogo **Plot** fornisce opzioni per il controllo dei complotti che sono in uscita per illustrare la potenza a due e tridimensionali mediante grafici di dimensioni di esempio ed effetto. La finestra di dialogo controlla anche i gradi di rotazione verticali e orizzontali per i grafici tridimensionali.

Grafico a due dimensioni

Confronto tra stima della potenza e dimensione del campione
Quando è abilitata, questa impostazione opzionale fornisce le opzioni per il controllo della potenza bidimensionale per dimensione del campione. L'impostazione è disabilitata per impostazione predefinita.

Intervallo dimensione campione
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione del campione.

Limite inferiore
Controlla il limite inferiore per la potenza bidimensionale per dimensione del campione. Il valore deve essere maggiore di 1 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la potenza bidimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Confronto tra stima della potenza e dimensione degli effetti
Per impostazione predefinita, questa impostazione opzionale è disabilitata. Quando è abilitata, il grafico visualizza nell'output. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , la gamma di tracciamento predefinita della dimensione di effetto utilizzata.

Intervallo della dimensione degli effetti
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori.

Limite inferiore
Controlla il limite inferiore per la potenza bidimensionale per effetto grafico ad effetto. Il valore deve essere maggiore o uguale a -5.0 e non può essere maggiore del valore **Limite superiore** .

Limite superiore
Controlla il limite superiore per la potenza bidimensionale per effetto grafico ad effetto. Il valore deve essere maggiore del valore **Limit. inferiore** e non può essere maggiore di 5.0.

Grafico a tre dimensioni

Confronto tra stima della potenza e
Fornisce opzioni per il controllo della potenza tridimensionale per dimensione del campione (x - asse) e dimensioni dell'effetto (asse y), le impostazioni di rotazione verticale / orizzontale e la gamma di tracciamento specificato dall'utente. Per impostazione predefinita, questa impostazione è disabilitata.

Dimensione degli effetti sull'asse x e dimensione del campione sull'asse y
L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (x - asse) e dimensioni dell'effetto (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Dimensione degli effetti sull'asse y e dimensione del campione sull'asse x
L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (asse y) e dimensione dell'effetto (asse x). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama

tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Intervallo dimensione campione
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione del campione.

Limite inferiore
Controlla il limite inferiore per la potenza tridimensionale per dimensione del campione. Il valore deve essere maggiore di 1 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la potenza tridimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Intervallo della dimensione degli effetti
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione dell'effetto.

Limite inferiore
Controlla il limite inferiore per la potenza tridimensionale per effetto grafico ad effetto. Il valore deve essere maggiore o uguale a -5.0 e non può essere maggiore del valore **Limite superiore** .

Limite superiore
Controlla il limite superiore per la potenza tridimensionale per effetto grafico ad effetto. Il valore deve essere maggiore del valore **Limit. inferiore** e non può essere maggiore di 5.0.

Analisi di potenza del test T a campioni indipendenti

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

Nell'analisi dei campioni indipendenti, i dati osservati contengono due campioni indipendenti. Si suppone che i dati in ogni campione seguano in modo indipendente e identicamente una normale distribuzione con una media e una varianza fissa, e richiama inferenza statistica sulla differenza dei due mezzi.

1. Dai menu, scegliere:

   Analizzare > **Analisi Alimentazione** > **Mezzi** > **Test di analisi indipendenti**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

   Per ulteriori informazioni, consultare “Analisi di alimentazione: Valori Grigi” a pagina 38.

   Immettere un **Rapporto dimensione gruppo** per specificare il rapporto tra le dimensioni del campione (il valore deve essere un valore singolo compreso tra 0.01 e 100.

4. Quando **Power** viene selezionato come metodo di test supposizione **Stima** , inserire i valori per specificare la dimensione del campione per **Dimensione di esempio per il gruppo 1 e il gruppo 2** per il confronto. I valori devono essere numeri interi superiori a 1.
5. Facoltativamente, selezionare un'opzione dal campo **Specificare** .

   Valori ipotizzati
   L'impostazione predefinita fornisce le impostazioni **mean Population mean** e **Population standard deviations** .

   Differenza media di popolazione
   Quando si specifica una sola media di popolazione, inserire il valore di differenza media della popolazione. Quando viene specificato un solo valore, denota la popolazione della differenza media di gruppo σ_d. Il valore deve essere un singolo numerico maggiore di 0.

   Popolazione media per gruppo 1 e gruppo 2
   Quando si specificano più mezzi di popolazione, inserire la popolazione media per i valori di gruppo 1 e di gruppo 2. Quando vengono specificati più valori, essi denotano la media di popolazione della differenza di gruppo σ_1 e σ_2. I valori devono essere un numero singolo superiore a 0.

   Le deviazioni standard della popolazione sono
   Specificare se le deviazioni standard della popolazione sono **Equal per due gruppi** o **Non uguale per due gruppi**.

   - Quando le deviazioni standard della popolazione sono uguali per due gruppi, inserire un valore per **Deviazione standard pooled** che denota σ, e presuppone che le due varianze di gruppo siano uguali, oppure $\sigma_1 = \sigma_2 = \sigma$.
   - Quando le deviazioni standard della popolazione non sono uguali per due gruppi, inserire valori per **deviazione standard per il gruppo 1 e il gruppo 2** che denotano σ_1 e σ_2.

     Nota: Quando i valori per **deviazione standard per il gruppo 1 e il gruppo 2** sono identici, vengono trattati come un valore unico.

   Dimensione effetto
   Stima la dimensione dell'effetto come input per la stima della potenza o della dimensione del campione. La dimensione dell'effetto definito **Valore** viene passata al passo intermedio nella procedura e calcola la potenza desiderata o la dimensione del campione.
6. Selezionare se il test è uno o due lati.

   Analisi non direzionale (bidirezionale)
   Quando viene selezionato, viene utilizzato un test a due lati. Questa è l'impostazione predefinita.

   Analisi direzionale (unilaterale)
   Quando selezionato, la potenza viene calcolata per un test unilaterale.
7. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.
8. Opzionalmente, clicca su **Plot** per specificare le impostazioni “Analisi di alimentazione di Independent - Samples T Test: Plot” a pagina 10 (output grafico, impostazioni della trama bidimensionale, impostazioni della trama tridimensionale e suggerimenti).

   Nota: Plot è disponibile solo quando **Power** viene selezionato come presupposto di prova.
9. Opzionalmente, clicca su **Precisione** per stimare la dimensione del campione in base agli intervalli di confidenza specificando i valori dell'intervallo di confidenza di mezza larghezza. Per ulteriori informazioni, consultare “Analisi di alimentazione: Precisione” a pagina 38.

Nota: Precision è disponibile solo quando **Dimensione di esempio** viene selezionata come metodo di test di test **Stima** , **Valori Hypothesize** è selezionato dall'elenco **Specificare** e **L'analisi non direzionale (a due lati)** viene selezionata come **Test Direction**.

Analisi di alimentazione di Independent - Samples T Test: Plot

La finestra di dialogo **Plot** fornisce opzioni per il controllo dei complotti che sono in uscita per illustrare la potenza a due e tridimensionali per rapporto di esempio, dimensione effetto o media differenza media. La finestra di dialogo controlla anche i gradi di rotazione verticali e orizzontali per i grafici tridimensionali.

Grafico a due dimensioni

Stima dell'alimentazione versus rapporto di dimensione del campione / stima di potenza versus dimensione del campione
Quando è abilitata, questa impostazione opzionale fornisce le opzioni per il controllo della potenza bidimensionale per il grafico del rapporto di dimensione del campione. Le impostazioni aggiuntive non sono disponibili quando sono stati specificati più valori di potenza (**Power stima versus sample size**). L'impostazione è disabilitata per impostazione predefinita.

Intervallo rapporto dimensione campione
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione del campione.

Limite inferiore
Controlla il limite inferiore per la potenza bidimensionale per il grafico del rapporto di dimensione del campione. Il valore deve essere compreso tra 0.01 e 100 e non può essere maggiore del valore **Limite superiore** .

Limite superiore
Controlla il limite superiore per la potenza bidimensionale per campione di rapporto di dimensione del campione. Il valore deve essere compreso tra 0.01 e 100 e deve essere maggiore del valore **Limiti inferiori** .

Confronto tra stima della potenza e dimensione degli effetti (o differenza media)
Per impostazione predefinita, questa impostazione opzionale è disabilitata. Quando è abilitata, il grafico visualizza nell'output. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione dell'effetto (o differenza media).

Intervallo della dimensione degli effetti (o differenza media)
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori.

Limite inferiore
Controlla il limite inferiore per la potenza bidimensionale per effetto grafico ad effetto. Il valore deve essere maggiore o uguale a -5.0 e non può essere maggiore del valore **Limite superiore** .

Limite superiore
Controlla il limite superiore per la potenza bidimensionale per effetto grafico ad effetto. Il valore deve essere maggiore del valore **Limit. inferiore** e non può essere maggiore di 5.0.

Grafico a tre dimensioni

Confronto tra stima della potenza e
Fornisce opzioni per il controllo della potenza tridimensionale per rapporto di dimensione del campione (x - asse) e dimensioni dell'effetto (asse y), le impostazioni di rotazione verticale / orizzontale e la gamma di tracciamento specificato dall'utente. Per impostazione predefinita, questa impostazione è disabilitata.

Dimensione effetto (o differenza media) su asse x e dimensione del campione su asse y
L'impostazione opzionale controlla la potenza tridimensionale per rapporto di dimensione del campione (x - asse) e dimensioni dell'effetto (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Dimensione effetto (o differenza media) su asse y e dimensione del campione su asse x
L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (asse y) e dimensione dell'effetto (asse x). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Intervallo rapporto dimensione campione
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione del campione.

Limite inferiore
Controlla il limite inferiore per la potenza tridimensionale per rapporto di dimensione del campione. Il valore deve essere compreso tra 0.01 e 100 e non può essere maggiore del valore **Limite superiore** .

Limite superiore
Controlla il limite superiore per la potenza tridimensionale per rapporto di dimensione del campione. Il valore deve essere compreso tra 0.01 e 100 e deve essere maggiore del valore **Limiti inferiori** .

Intervallo della dimensione degli effetti (o differenza media)
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione dell'effetto.

Limite inferiore
Controlla il limite inferiore per la potenza tridimensionale per effetto grafico ad effetto. Il valore deve essere maggiore o uguale a -5.0 e non può essere maggiore del valore **Limite superiore** .

Limite superiore
Controlla il limite superiore per la potenza tridimensionale per effetto grafico ad effetto. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Analisi di potenza di ANOVA a una via

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

L'analisi della varianza (ANOVA) è un metodo statistico di stima dei mezzi di diverse popolazioni che spesso si presume siano normalmente distribuiti. La One-way ANOVA, un tipo comune di ANOVA, è un'estensione del test a due sample *t*. La procedura fornisce approcci per stimare la potenza per due tipi di ipotesi per confrontare i mezzi di gruppo multipli, il test complessivo e il test con contrasti specificati. Il test over si concentra sull'ipotesi nulla che tutti i mezzi di gruppo siano uguali. Il test con contrasti specificati rompe le ipotesi di ANOVA complessive in pezzi più piccoli ma più descrivibili e utili dei mezzi.

1. Dai menu, scegliere:

   Analizzare > **Analisi Alimentazione** > **Mezzi** > **One-way ANOVA**

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

   Per ulteriori informazioni, consultare "Analisi di alimentazione: Valori Grigi" a pagina 38.
4. Facoltativamente, selezionare un'opzione dal campo **Specificare** .

   Valori ipotizzati
   L'impostazione predefinita fornisce l'impostazione **Pooled Standard deviazione standard** . Inserire un valore numerico unico superiore a 0. Il valore predefinito è 1.

   Dimensione effetto: f di Cohen
   Il f di Cohen viene utilizzato per stimare la dimensione dell'effetto come input per la stima della potenza o della dimensione del campione. La dimensione dell'effetto definito **Valore** viene passata al passo intermedio nella procedura e calcola la potenza desiderata o la dimensione del campione.

   Dimensione effetto: Eta quadratico
   Et - quadrato (η^2) viene utilizzato per stimare la dimensione dell'effetto come input per la stima della potenza o della dimensione del campione. La dimensione dell'effetto definito **Valore** viene passata al passo intermedio nella procedura e calcola la potenza desiderata o la dimensione del campione.
5. Quando **Power** viene selezionata come impostazione di test supposizione **Stima** , specificare i valori **Dimensioni di gruppo** e **Gruppo di gruppi** . Devono essere specificati almeno due valori di dimensione del gruppo e ognuno dei valori di dimensione del gruppo deve essere maggiore o uguale a 2 (ma non può superare i 2.147, 483, 647). Devono essere specificati anche almeno due valori medi di gruppo (il numero di valori specificati deve egualare i valori di dimensione del gruppo).

   Quando **Dimensione di esempio** viene selezionata come impostazione dell'assunzione di test **Stima** , specificare i valori di **Gruppo pesi** (opzionali) e **Group.** . I pesi di gruppo assegnano i pesi di dimensione del gruppo quando **Power** viene selezionato.

   Nota: Le impostazioni dei **pesi del gruppo** vengono ignorate quando vengono specificati i valori **Dimensioni di gruppo** .

   Clicca su **Aggiungi** per includere più **Dimensioni di gruppo**, **pesi di gruppo**e **I mezzi di gruppo** . Clicca su **Elimina** per rimuovere i valori esistenti.
6. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.
7. Opzionalmente, clicca su **Contrast** per specificare le impostazioni "Analisi alimentazione di One-way ANOVA: contrasto" a pagina 13 (test di contrasto, differenze pairwise, dimensioni degli effetti e stima la dimensione del campione in base agli intervalli di confidenza).
8. Opzionalmente, clicca su **Plot** per specificare le impostazioni "Analisi di alimentazione di One-way ANOVA: Plot" a pagina 13 (output grafico, impostazioni della trama bidimensionale, impostazioni della trama tridimensionale e suggerimenti).

   Nota: Plot è disponibile solo quando vengono specificati i valori **Dimensioni di gruppo** e **Power** viene selezionata come impostazione **Stima** .
9. Opzionalmente, clicca su **Precisione** per stimare la dimensione del campione in base agli intervalli di confidenza specificando i valori dell'intervallo di confidenza di mezza larghezza. Per ulteriori informazioni, consultare "Analisi di alimentazione: Precisione" a pagina 38.

Nota: Precision è disponibile solo quando **Dimensione di esempio** viene selezionata come metodo di test supposizione **Stima** .

Analisi alimentazione di One-way ANOVA: contrasto

La finestra di dialogo **Contrast** fornisce opzioni per specificare il contrasto, il coefficiente, la dimensione dell'effetto, le differenze pairwise e l'intervallo di confidenza con impostazioni di mezza larghezza per l'analisi di Power Analysis di One-way ANOVA.

Test del contrasto

Test con contrasti lineari
Quando è abilitata, le impostazioni del contrasto e del coefficiente sono disponibili.

Direzione del test

Analisi non direzionale (bidirezionale)
Quando viene selezionato, viene utilizzato un test a due lati. Questa è l'impostazione predefinita.

Analisi direzionale (unilaterale)
Quando selezionato, la potenza viene calcolata per un test unilaterale.

Coefficienti
Utilizzare la tabella per specificare i coefficienti di contrasto e richiedere il test di contrasto. I valori della tabella sono facoltativi. Il numero di valori specificati deve essere uguale ai valori specificati per **Dimensioni di gruppo** e **Gruppo significa**. La somma per tutti i valori specificati deve essere uguale a 0, altrimenti l'ultimo valore verrà automaticamente regolato.

Dimensione degli effetti
Specificare la dimensione dell'effetto del test di contrasto misurato da δ. È necessario specificare un valore numerico unico. Quando la **Direzione dei test** è impostata su **Analisi direzionale (unilaterale)**, il valore specificato deve essere ≥ 0. Quando **Power** viene specificato come l'assunzione di test **Stima** , il valore specificato non può essere 0.

Differenze pairwise

Stimare la potenza dei test per le differenze pairwise
Controlla se stimare o meno il potere di verifica per le differenze pairwise. Di default, l'impostazione opzionale è disabilitata, che sopprime l'output per le differenze pairwise.

Correzione di Bonferroni
Utilizza la correzione Bonferroni nella stima del potere delle differenze pairwise. Questa è l'impostazione predefinita.

Sidak corretta
Utilizza la correzione di Sidak per stimare il potere delle differenze pairwise.

Differenza meno significativa (LSD)
Utilizza la correzione LSD nella stima del potere delle differenze pairwise.

Specificare la metà - larghezza dell'intervallo di confidenza
Stima la dimensione del campione in base al valore di mezza larghezza di confidenza. Inserire un valore nella gamma 0 - 1. Per il test binomiale a campione singolo, il valore deve essere compreso nell'intervallo 0 - 0.5.

Nota: I valori duplicati vengono ignorati.

- Clicca su **Aggiungi** per aggiungere il valore di mezza larghezza specificato all'elenco.
- Evidenziare un valore di mezza larghezza esistente e fare clic su **Modifica** per aggiornare il valore.
- Evidenziare un valore di mezza larghezza esistente e fare clic su **Rimuovi** per rimuovere il valore dall'elenco.

Analisi di alimentazione di One-way ANOVA: Plot

La finestra di dialogo **Plot** controlla i complotti che sono in uscita per illustrare la potenza a due e tridimensionali mediante grafici di dimensioni di esempio ed effetto. La finestra di dialogo controlla anche i gradi di rotazione verticali e orizzontali per i grafici tridimensionali.

Grafico a due dimensioni

Confronto tra stima della potenza e dimensione del campione totale
Quando è abilitata, questa impostazione opzionale fornisce opzioni per il controllo della potenza bidimensionale per dimensione totale del campione. L'impostazione è disabilitata per impostazione predefinita.

Intervallo dimensione campione totale
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione totale del campione.

Limite inferiore
Controlla il limite inferiore per la potenza bidimensionale per total sample size chart. Il valore deve essere superiore o uguale a:

- 2 x il numero di numeri interi specificati per **Dimensioni di gruppo**
- 2 x la somma degli interi specificati per **Dimensioni di gruppo** / dal valore intero più piccolo per **Dimensioni di gruppo**

Il valore non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la potenza bidimensionale per total sample size chart. Il valore deve essere inferiore o uguale a:

- 5000 / per il valore intero più grande specificato per **Dimensioni di gruppo** x la somma degli interi specificati per **Dimensioni di gruppo**

Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 2147483647.

Confronto tra stima della potenza e deviazione standard raggruppata
Per impostazione predefinita, questa impostazione opzionale è disabilitata. L'impostazione controlla la potenza bidimensionale per diagramma di deviazione standard pooled. Quando è abilitata, il grafico visualizza nell'output. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzato il range di tracciamento predefinito della deviazione standard pooled.

Nota:

La trama è disabilitata quando i valori specificati **Group significa** sono tutti uguali.

Intervallo della deviazione standard raggruppata
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori.

Limite inferiore
Controlla il limite inferiore per la potenza bidimensionale tramite diagramma di deviazione standard pooled. Il valore deve essere maggiore di 0 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la potenza bidimensionale tramite diagramma di deviazione standard pooled. Il valore deve essere maggiore del valore **Bassa bound** .

Grafico a tre dimensioni

Confronto tra stima della potenza e
Fornisce opzioni per il controllo della potenza tridimensionale per dimensione totale del campione (x - asse) e dimensioni dell'effetto (asse y), le impostazioni di rotazione verticale e orizzontale e la gamma di tracciamento specificato dall'utente. Per impostazione predefinita, questa impostazione è disabilitata.

Nota:

La trama è disabilitata quando i valori specificati **Group significa** sono tutti uguali.

Deviazione standard raggruppata sull'asse x e dimensione totale del campione sull'asse y
L'impostazione opzionale controlla la potenza tridimensionale per dimensione totale del campione (x - asse) e grafico di deviazione standard (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Deviazione standard raggruppata sull'asse y e dimensione totale del campione sull'asse x
L'impostazione opzionale controlla la potenza tridimensionale per dimensione totale del campione (asse y) e grafico a deviazione standard (asse x). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Intervallo dimensione campione totale
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione del campione.

Limite inferiore
Controlla il limite inferiore per la potenza tridimensionale per total sample size chart. Il valore deve essere maggiore di 0 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la potenza tridimensionale per total sample size chart. Il valore deve essere maggiore del valore **Bassa bound** .

Intervallo della deviazione standard raggruppata
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione dell'effetto.

Limite inferiore
Controlla il limite inferiore per la potenza tridimensionale mediante diagramma di deviazione standard pooled. Il valore deve essere maggiore di 0 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la potenza tridimensionale mediante diagramma di deviazione standard pooled. Il valore deve essere maggiore del valore **Bassa bound** .

Confronto tra stima della potenza e dimensione del campione
Fornisce opzioni per il controllo della potenza tridimensionale per dimensione e l'intervallo di tracciamento specificato dall'utente e dimensione di effetto. Per impostazione predefinita, questa impostazione è disabilitata.

Nota:

La trama è disabilitata quando i valori specificati **Group significa** sono tutti uguali.

Deviazione standard raggruppata sull'asse x e dimensione totale del campione sull'asse y
L'impostazione opzionale controlla la potenza tridimensionale per dimensione totale del campione (x - asse) e dimensioni dell'effetto (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Deviazione standard raggruppata sull'asse y e dimensione totale del campione sull'asse x
L'impostazione opzionale controlla la potenza tridimensionale per dimensione totale del campione (asse y) e dimensione dell'effetto (asse x). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Intervallo dimensione campione totale
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione del campione.

Limite inferiore
Controlla il limite inferiore per la potenza tridimensionale per total sample size chart. Il valore deve essere maggiore di 0 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la potenza tridimensionale per total sample size chart. Il valore deve essere maggiore del valore **Bassa bound** .

Intervallo della dimensione degli effetti
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di tracciamento predefinita della dimensione dell'effetto.

Limite inferiore
Controlla il limite inferiore per la potenza tridimensionale per total sample size chart. Il valore deve essere maggiore di 0 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la potenza tridimensionale per total sample size chart. Il valore deve essere maggiore del valore **Bassa bound** .

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Proporzioni

Le seguenti funzioni statistiche sono incluse in IBM SPSS Statistica Base Edition.

Analisi di potenza del test binomiale a un campione

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

La distribuzione binomiale si basa su una sequenza di prove di Bernoulli. Può essere utilizzato per modellare esperimenti, tra cui un numero fisso di prove totali, che si presume siano indipendenti l'uno dall'altro. Ogni processo porta a un risultato dicotomico, con la stessa probabilità per un risultato di successo.

Il test binomiale di un campione effettua inferenza statistica sul parametro di proporzione confrontandolo con un valore ipotizzato. I metodi per stimare la potenza per un tale test sono la normale approssimazione o l'enumerazione binomiale.

1. Dai menu, scegliere:

   Analizzare > **Analisi Alimentazione** > **Proporzioni** > **Test binario di esempio**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure

selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

Per ulteriori informazioni, consultare "Analisi di alimentazione: Valori Grigi" a pagina 38.

4. Quando **Power** viene selezionata come impostazione di prova **Stima** , inserire l'appropriato valore **Numero totale di prove** . Il valore deve essere un intero maggiore di, o uguale a, 1.
5. Inserire un valore che specifica il valore di ipotesi alternativo del parametro di proporzione nel campo **Percentuale di popolazione** . Il valore deve essere un singolo numerico.

   Nota: Quando viene specificato un valore **Power** , il valore **Percentuale di Popolazione** non può essere uguale al **Valore Null**.
6. Facoltativamente, inserire un valore che specifica il valore di ipotesi nullo del parametro proporzione da verificare nel campo **Valore Null** . Il valore deve essere un singolo numerico nella gamma 0 - 1. Il valore predefinito è 0.50.
7. Selezionare un metodo per stimare la potenza.

   Approssimazione normale
   Abilita l'approssimazione normale. Questa è l'impostazione predefinita.

   Applica correzione della continuità
   Controllare se la correzione di continuità viene utilizzata o meno per il metodo di approssimazione normale.

   Enumerazione binomiale
   Abilita il metodo di enumerazione binomiale. Opzionalmente, utilizzare il campo **Limite di tempo** per specificare il numero massimo di minuti consentito per stimare la dimensione del campione. Quando viene raggiunto il limite di tempo, l'analisi viene terminata e viene visualizzato un messaggio di avviso. Quando specificato, il valore deve essere un intero intero positivo per denotare il numero di minuti. L'impostazione predefinita è 5 minutes.

   Nota: L'assunzione di stima dell'alimentazione selezionata non ha alcun effetto quando il valore **Numero totale di prove** supera il 500.
8. Selezionare se il test è uno o due lati.

   Analisi non direzionale (bidirezionale)
   Quando viene selezionato, viene utilizzato un test a due lati. Questa è l'impostazione predefinita.

   Analisi direzionale (unilaterale)
   Quando selezionato, la potenza viene calcolata per un test unilaterale.
9. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.
10. Opzionalmente, clicca su **Plot** per specificare le impostazioni "Analisi alimentazione di One - Esempio Binomiale: Plot" a pagina 17 (output grafico, impostazioni della trama bidimensionale e impostazioni della trama tridimensionale).

    Nota: Plot è disponibile solo quando **Power** viene selezionato come l'assunzione di test **Stima** e **L'enumerazione binomiale** non è selezionata.
11. Opzionalmente, clicca su **Precisione** per stimare la dimensione del campione in base agli intervalli di confidenza specificando i valori dell'intervallo di confidenza di mezza larghezza. Per ulteriori informazioni, consultare "Analisi di alimentazione: Precisione" a pagina 38.

Nota: Precision è disponibile solo quando **Dimensione di esempio** viene selezionata come metodo di test supposizione **Stima** e l'analisi **Non direzionale (a due lati)** viene selezionata come **Test Direction**.

Analisi alimentazione di One - Esempio Binomiale: Plot

La finestra di dialogo **Plot** fornisce opzioni per il controllo dei complotti che sono in uscita per illustrare la potenza a due e tridimensionali per grafici. La finestra di dialogo controlla anche i gradi di rotazione verticali e orizzontali per i grafici tridimensionali.

Grafico a due dimensioni

Fornisce opzioni per il controllo della stima di potenza bidimensionale versus grafici. Per impostazione predefinita, questa impostazione è disabilitata.

Confronto tra stima della potenza e valore dell'ipotesi nulla

Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per grafico a valore nullo. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e valore dell'ipotesi alternativa

Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per grafico di valore alternativo. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e la differenza tra valori ipotizzati

Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per differenza tra il grafico dei valori ipotizzati. L'impostazione è disabilitata per impostazione predefinita.

Confronto tra stima della potenza e numero totale di prove

Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per numero totale di trial grafico. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Intervallo grafico del numero totale di prove

Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore

Controlla il limite inferiore per la stima di potenza bidimensionale versus numero totale di trial grafico. Il valore deve essere maggiore di 0 e non può essere maggiore del valore **Upper bound** .

Limite superiore

Controlla il limite superiore per la stima di potenza bidimensionale versus numero totale di trial grafico. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Grafico a tre dimensioni

Fornisce opzioni per il controllo della stima di potenza tridimensionale versus grafici. Per impostazione predefinita, questa impostazione è disabilitata.

Confronto tra stima della potenza e numero totale di prove

Quando selezionata, questa impostazione abilita le seguenti opzioni.

sull'asse x e la differenza tra i valori ipotizzati su asse y

L'impostazione opzionale controlla la potenza tridimensionale per numero totale di prove (x -asse) e differenza tra i valori ipotizzati (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

sull'asse y e la differenza tra i valori ipotizzati su asse x

L'impostazione opzionale controlla la potenza tridimensionale per numero totale di prove (asse y) e differenza tra i valori ipotizzati (x -asse). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Intervallo grafico del numero totale di prove

Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore

Controlla il limite inferiore per la stima di potenza tridimensionale rispetto al numero totale di trial grafico. Il valore deve essere maggiore di 0 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima di potenza tridimensionale contro il numero totale di trial grafico. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Confronto tra stima della potenza e valore dell'ipotesi nulla
Quando selezionata, questa impostazione abilita le seguenti opzioni.

su asse x e valore dell'ipotesi alternativa su asse y
L'impostazione opzionale controlla la potenza tridimensionale per grafico nullo (x -asse) e valore alternativo (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

su asse y e valore dell'ipotesi alternativa su asse x
L'impostazione opzionale controlla la potenza tridimensionale da parte di un grafico nullo (asse y) e valore alternativo (x -asse) .. Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Analisi di potenza del test binomiale a campioni correlati

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

La distribuzione binomiale si basa su una sequenza di prove di Bernoulli. Può essere utilizzato per modellare esperimenti, tra cui un numero fisso di prove totali, che si presume siano indipendenti l'uno dall'altro. Ogni processo porta a un risultato dicotomico, con la stessa probabilità per un risultato di successo.

Il binomio campione correlato stima la potenza del test McNemar's per confrontare due parametri di proporzione basati su una coppia di soggetti abbinati campionati da due popolazioni binomiali correlate.

1. Dai menu, scegliere:

   Analizzare > **Analisi di alimentazione** > **proporzioni** > **Test binario di esempio correlato**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

   Per ulteriori informazioni, consultare “Analisi di alimentazione: Valori Grigi” a pagina 38.

4. Quando **Power** viene selezionato come metodo di test supposizione **Stima** , inserire l'apposito valore **Numero totale di coppie** .
5. Selezionare per specificare i valori di verifica per **proporzioni** o **conta**.
   - Quando vengono selezionate **proporzioni** , inserire i valori nei campi **Proporzione 1** e **Proporzione 2** . I valori devono essere compresi nella gamma 0 - 1.
   - Quando **conta** viene selezionato, inserire valori nei campi **Conteggio 1** e **Conteggio 2** . I valori devono essere compresi tra 0 e valore specificati per **Numero totale di coppie**.

   Note di proporzioni:

   - **Proporzioni** è l'unica opzione disponibile quando viene specificato un valore **Alimentazione** .
   - Quando **I valori di test sono marginali** non è selezionato: 0 < **Proporzione 1** + **Proporzione 2** ≤ 1
   - Quando **I valori di test sono marginali** viene selezionato:
     - **Proporzione 1** * **Proporzione 2** > 0
     - **Proporzione 1** < 1
     - **Proporzione 2** < 1
     - I valori per **Proporzione 1** e **Proporzione 2** non possono essere uguali.

   Note sui conteggi:

   - Le impostazioni **conteggi** sono disponibili solo quando **Power** viene selezionato come impostazione di prova **Stima** .
   - Quando **I valori di test sono marginali** non è selezionato: 0 < **Conteggio 1** + **Conteggio 2** ≤ **Numero totale di coppie**
   - Quando **I valori di test sono marginali** viene selezionato:
     - **Conteggio 1** * **Conteggio 2** > 0
     - **Conteggio 1** < **Numero totale di coppie**
     - **Conteggio 2** < **Numero totale di coppie**
6. È possibile selezionare facoltativamente **I valori di test sono marginali** per controllare se le proporzioni o i conteggi specificati sono marginali. Quando **I valori di test sono marginali** è abilitato, è necessario specificare un valore **Correlazione tra coppie corrispondenti** . Il valore deve essere un singolo valore compreso tra -1 e 1.
7. Selezionare un metodo per stimare la potenza.

   Approssimazione normale
   Abilita l'approssimazione normale. Questa è l'impostazione predefinita.

   Enumerazione binomiale
   Abilita il metodo di enumerazione binomiale. Opzionalmente, utilizzare il campo **Limite di tempo** per specificare il numero massimo di minuti consentito per stimare la dimensione del campione. Quando viene raggiunto il limite di tempo, l'analisi viene terminata e viene visualizzato un messaggio di avviso. Quando specificato, il valore deve essere un intero intero positivo per denotare il numero di minuti. L'impostazione predefinita è 5 minutes.
8. Selezionare se il test è uno o due lati.

   Analisi non direzionale (bidirezionale)
   Quando viene selezionato, viene utilizzato un test a due lati. Questa è l'impostazione predefinita.

   Analisi direzionale (unilaterale)
   Quando selezionato, la potenza viene calcolata per un test unilaterale.
9. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.

10. Opzionalmente, clicca su **Plot** per specificare le impostazioni “Analisi di potenza del test binomiale a campioni correlati” a pagina 19 (output grafico, impostazioni della trama bidimensionale e impostazioni della trama tridimensionale).

    Nota: Plot è disponibile solo quando **Power** viene selezionato come l'assunzione di test **Stima** e **L'enumerazione binomiale** non è selezionata.
11. Opzionalmente, clicca su **Precisione** per stimare la dimensione del campione in base agli intervalli di confidenza specificando i valori dell'intervallo di confidenza di mezza larghezza. Per ulteriori informazioni, consultare “Analisi di alimentazione: Precisione” a pagina 38.

Nota: Precision è disponibile solo quando **Dimensione di esempio** viene selezionata come metodo di test supposizione **Stima** e l'analisi **Non direzionale (a due lati)** viene selezionata come **Test Direction**.

Analisi di potenza di Correlato - Esempio Binomiale: Plot

La finestra di dialogo **Plot** controlla i complotti che sono in uscita per illustrare i due e i grafici di stima della potenza tridimensionali. La finestra di dialogo controlla anche i gradi di rotazione verticali e orizzontali per i grafici tridimensionali.

Grafico a due dimensioni
Fornisce opzioni per il controllo della stima di potenza bidimensionale versus grafici. Per impostazione predefinita, questa impostazione è disabilitata.

Confronto tra stima della potenza e numero totale di coppie
Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per numero totale di coppie. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Intervallo grafico del numero totale di coppie
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per la stima della potenza bidimensionale contro il numero totale di coppie. Il valore deve essere maggiore di 1 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima della potenza bidimensionale contro il numero totale di coppie. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 2500.

Confronto tra stima della potenza e differenza di rischio
Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per differenza di rischio grafico. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e rapporto di rischio
Quando abilitato, questa impostazione opzionale controlla la potenza bidimensionale per rapporto di rapporto di rischio. L'impostazione è disabilitata per impostazione predefinita.

Intervallo grafico del rapporto di rischio
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per la stima della potenza bidimensionale contro il rapporto di rischio. Il valore non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima dell'alimentazione bidimensionale versus rapporto di rapporto di rischio. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 10.

Confronto tra stima della potenza e rapporto odd
Quando è abilitata, questa impostazione opzionale controlla la potenza bidimensionale per odds ratio chart. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Intervallo grafico del rapporto odd
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per il grafico di stima dell'alimentazione bidimensionale versus odds ratio. Il valore non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima dell'alimentazione bidimensionale versus odds ratio chart. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 10.

Confronto tra stima della potenza e correlazione tra coppie corrispondenti
Controlla la potenza stimata rispetto alla correlazione tra il grafico delle coppie corrispondenti. Il grafico viene creato solo quando vengono specificate le proporzioni marginali o i conteggi (piuttosto che le proporzioni discordanti).

Grafico a tre dimensioni
Fornisce opzioni per il controllo della stima di potenza tridimensionale versus grafici. Per impostazione predefinita, questa impostazione è disabilitata.

Confronto tra stima della potenza e proporzioni discordanti
Quando abilitata, questa impostazione opzionale controlla la potenza tridimensionale per discordante proporzioni grafico. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e proporzioni marginali
Quando è abilitata, questa impostazione opzionale controlla la potenza tridimensionale per proporzioni marginali. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Nota: Questa impostazione è disponibile solo quando **I valori di test sono marginali** viene selezionato.

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Analisi di potenza del test binomiale a campioni indipendenti

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

La distribuzione binomiale si basa su una sequenza di prove di Bernoulli. Può essere utilizzato per modellare quegli esperimenti tra cui un numero fisso di prove totali che si presume siano indipendenti l'una dall'altra. Ogni processo porta a un risultato dicotomico, con la stessa probabilità per un risultato "di successo". Il test binomiale a campione indipendente confronta due parametri di proporzione indipendenti.

1. Dai menu, scegliere:

   Analizzare > **Analisi di alimentazione** > **proporzioni** > **Test binomiale dei campioni indipendenti**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

   Per ulteriori informazioni, consultare “Analisi di alimentazione: Valori Grigi” a pagina 38.

   Facoltativamente, specificare un valore **Dimensione del gruppo** . Il valore predefinito è 1.
4. Quando **Power** viene selezionata come impostazione di test supposizione **Stima** , inserire i valori per specificare il **Numero totale di prove per il gruppo 1 e il gruppo 2**. I valori devono essere numeri interi superiori a 1.
5. Specificare i parametri di proporzione per i due gruppi. Entrambi i valori devono essere compresi nella gamma 0 - 1.

   Nota: I due valori non possono essere gli stessi quando viene specificato un valore **Alimentazione** .
6. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.
7. Selezionare il metodo di prova desiderato:

   Test del chi-quadrato
   Stima la potenza in base al test chi quadrato di Pearson. Questa è l'impostazione predefinita.

   La deviazione standard è raggruppata
   Questa impostazione opzionale controlla se la stima della deviazione standard è pooled o scaricata. L'impostazione è abilitata per impostazione predefinita.

   Applica correzione della continuità
   Questa impostazione opzionale controlla se viene utilizzata o meno la correzione di continuità. L'impostazione è disabilitata per impostazione predefinita.

   Test T
   Stima la potenza in base al t - test di Student.

   La deviazione standard è raggruppata
   Questa impostazione opzionale controlla se la stima della deviazione standard è pooled o scaricata. L'impostazione è abilitata per impostazione predefinita.

   Test rapporto di verosimiglianza
   Stima la potenza in base al test del rapporto di verosimiglie.

   test esatto di Fisher
   Stima il potere basato sul test esatto di Fisher.

   Note:

   - In alcuni casi, il test esatto di Fisher potrebbe richiedere un tempo prolungato di tempo per completare.
   - Tutti i tracciati sono bloccati quando viene selezionato il test esatto di Fisher.
8. Selezionare un metodo per stimare la potenza.

Approssimazione normale
Abilita l'approssimazione normale. Questa è l'impostazione predefinita.

Enumerazione binomiale
Abilita il metodo di enumerazione binomiale. Opzionalmente, utilizzare il campo **Limite di tempo** per specificare il numero massimo di minuti consentito per stimare la dimensione del campione. Quando viene raggiunto il limite di tempo, l'analisi viene terminata e viene visualizzato un messaggio di avviso. Quando specificato, il valore deve essere un intero intero positivo per denotare il numero di minuti. L'impostazione predefinita è 5 minutes.

9. Selezionare se il test è uno o due lati.

Analisi non direzionale (bidirezionale)
Quando viene selezionato, viene utilizzato un test a due lati. Questa è l'impostazione predefinita.

Analisi direzionale (unilaterale)
Quando selezionato, la potenza viene calcolata per un test unilaterale.

10. È possibile opzionalmente fare clic su **Plot** per specificare le impostazioni "Analisi di alimentazione di Independent - Samples Binomial Test: Plot" a pagina 24 (output grafico, impostazioni della trama bidimensionale e impostazioni della trama tridimensionale).

Nota: Plot è disponibile solo quando **Power** viene selezionato come l'assunzione di test **Stima** e **L'enumerazione binomiale** non è selezionata.

11. Opzionalmente, clicca su **Precisione** per stimare la dimensione del campione in base agli intervalli di confidenza specificando i valori dell'intervallo di confidenza di mezza larghezza. Per ulteriori informazioni, consultare "Analisi di alimentazione: Precisione" a pagina 38.

Nota: Precision è disponibile solo quando **Dimensione di esempio** viene selezionata come metodo di test supposizione **Stima** e l'analisi **Non direzionale (a due lati)** viene selezionata come **Test Direction**.

Analisi di alimentazione di Independent - Samples Binomial Test: Plot

La finestra di dialogo **Plot** fornisce opzioni per il controllo dei complotti che sono in uscita per illustrare i due e i grafici di stima della potenza tridimensionali. La finestra di dialogo controlla anche i gradi di rotazione verticali e orizzontali per i grafici tridimensionali.

Grafico a due dimensioni
Fornisce opzioni per il controllo della stima di potenza bidimensionale versus grafici. Per impostazione predefinita, questa impostazione è disabilitata.

Stima dell'alimentazione versus rapporto di dimensione del gruppo / stima di potenza versus dimensione di gruppo
Quando è abilitata, questa impostazione opzionale controlla la potenza bidimensionale per rapporto di dimensione del gruppo. Quando selezionata, questa impostazione visualizza il grafico. Le impostazioni aggiuntive non sono disponibili quando sono stati specificati più valori di potenza (**Power stima versus group size**). L'impostazione è disabilitata per impostazione predefinita.

Intervallo grafico del rapporto dimensione gruppo
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per la stima della potenza bidimensionale contro il numero totale di coppie. Il valore deve essere maggiore di .01 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima della potenza bidimensionale contro il numero totale di coppie. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 100.

Confronto tra stima della potenza e differenza di rischio
Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per differenza di rischio grafico. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e rapporto di rischio
Quando abilitato, questa impostazione opzionale controlla la potenza bidimensionale per rapporto di rapporto di rischio. L'impostazione è disabilitata per impostazione predefinita.

Intervallo grafico del rapporto di rischio
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per la stima della potenza bidimensionale contro il rapporto di rischio. Il valore non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima dell'alimentazione bidimensionale versus rapporto di rapporto di rischio. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 10.

Confronto tra stima della potenza e rapporto odd
Quando è abilitata, questa impostazione opzionale controlla la potenza bidimensionale per odds ratio chart. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Intervallo grafico del rapporto odd
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per il grafico di stima dell'alimentazione bidimensionale versus odds ratio. Il valore non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima dell'alimentazione bidimensionale versus odds ratio chart. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 10.

Grafico a tre dimensioni
Fornisce opzioni per il controllo della stima di potenza tridimensionale versus grafici. Per impostazione predefinita, questa impostazione è disabilitata.

Stima dell'alimentazione versus proporzioni
Quando selezionata, questa impostazione opzionale fornisce la seguente potenza per proporzione di opzioni:

proporzione del gruppo 1 su asse x e proporzione del gruppo 2 su asse y
Controlla la potenza tridimensionale per proporzione di gruppo 1 (x -asse) e proporzione del grafico di gruppo 2 (asse y). L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

proporzione del gruppo 1 su asse y e proporzione del gruppo 2 su asse x
Controlla la potenza tridimensionale per proporzione di gruppo 2 (x -asse) e proporzione del grafico di gruppo 1 (asse y). L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Stima dell'alimentazione versus dimensioni di gruppo
Quando selezionata, questa impostazione opzionale fornisce la seguente potenza per opzioni di dimensioni di gruppo:

dimensione del gruppo 1 su asse x e dimensione del gruppo 2 su asse y
Controlla la potenza tridimensionale per numero di prove nel gruppo 1 (x -asse) e numero di prove nel grafico di gruppo 2 (asse y). L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

dimensione del gruppo 1 su asse y e dimensione del gruppo 2 su asse x
Controlla la potenza tridimensionale per numero di prove nel gruppo 2 (x -asse) e numero di prove nel grafico di gruppo 1 (asse y). L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Intervallo grafico di dimensioni del gruppo 1
Quando selezionate, sono disponibili le opzioni di collegamento inferiore e superiore per la gamma di trama del gruppo 1. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per il grafico di stima dell'alimentazione bidimensionale versus odds ratio. Il valore deve essere maggiore o uguale a 2 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima dell'alimentazione bidimensionale versus odds ratio chart. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 2500.

Intervallo grafico di dimensioni del gruppo 2
Quando selezionate, sono disponibili le opzioni di collegamento inferiore e superiore per la gamma di trama del gruppo 2. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per il grafico di stima dell'alimentazione bidimensionale versus odds ratio. Il valore deve essere maggiore o uguale a 2 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima dell'alimentazione bidimensionale versus odds ratio chart. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 2500.

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Correlazioni

Le seguenti funzioni statistiche sono incluse in IBM SPSS Statistica Base Edition.

Analisi di potenza del test di correlazione di Pearson a un campione

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera

ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

Il coefficiente di correlazione del momento prodotto da Pearson misura la forza dell'associazione lineare tra due variabili casuali su scala che si ipotizza di seguire una distribuzione normale bivariata. Per convenzione, si tratta di una quantità adimensionale e si ottiene standardizzando la covarianza tra due variabili continue, quindi compresa tra -1 e 1.

Il test utilizza il metodo asintotico di Fisher per stimare la potenza per la correlazione di Pearson a campione.

1. Dai menu, scegliere:

   Analizzare > **Analisi Alimentazione** > **Correlazioni** > **Pearson Product - Moment**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

   Per ulteriori informazioni, consultare “Analisi di alimentazione: Valori Grigi” a pagina 38.
4. Quando **Power** viene selezionata come impostazione di prova **Stima** , inserire l'apposito valore **Dimensione di esempio in coppie** . Il valore deve essere un intero intero maggiore di 3.
5. Inserire un valore che specifica il valore di ipotesi alternativo del parametro di correlazione nel campo **Pearson correlazione parametro** . Il valore deve essere un numero singolo compreso tra -1 e 1.

   Nota: Quando si specifica **Power**, il valore **del parametro di correlazione di Pearson** non può essere -1 o 1 e non può essere uguale al **valore Null**.
6. Facoltativamente, inserire un valore che specifica il valore di ipotesi nullo del parametro di correlazione da verificare nel campo **Valore Null** . Il valore deve essere un numero singolo compreso tra -1 e 1. Il valore predefinito è 0.

   Nota: Quando si specifica **Power**, il **valore Null** non può essere -1 o 1.
7. Opzionalmente, selezionare **Utilizza bias - correzione formiola nella stima dell'alimentazione** per specificare se la regolazione del bias è coinvolta o ignorata. L'impostazione è abilitata per impostazione predefinita, che include il termine di regolazione della bias nella stima dell'alimentazione. Quando l'impostazione non viene selezionata, il termine di regolazione della bias viene ignorato.
8. Selezionare se il test è uno o due lati.

   Analisi non direzionale (bidirezionale)
   Quando viene selezionato, viene utilizzato un test a due lati. Questa è l'impostazione predefinita.

   Analisi direzionale (unilaterale)
   Quando selezionato, la potenza viene calcolata per un test unilaterale.
9. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.
10. È possibile opzionalmente fare clic su **Plot** per specificare le impostazioni “Analisi di alimentazione di One - Sample Pearson Correlazione: Plot” a pagina 28 (output grafico, impostazioni della trama bidimensionale e impostazioni della trama tridimensionale).

    Nota: Plot è disponibile solo quando **Power** viene selezionato come presupposto di prova.
11. Opzionalmente, clicca su **Precisione** per stimare la dimensione del campione in base agli intervalli di confidenza specificando i valori dell'intervallo di confidenza di mezza larghezza. Per ulteriori informazioni, consultare “Analisi di alimentazione: Precisione” a pagina 38.

Nota: Precision è disponibile solo quando **Dimensione di esempio** viene selezionata come metodo di test supposizione **Stima** e l'analisi **Non direzionale (a due lati)** viene selezionata come **Test Direction**.

Analisi di alimentazione di One - Sample Pearson Correlazione: Plot

La finestra di dialogo **Plot** fornisce opzioni per il controllo dei complotti che sono in uscita per illustrare la potenza a due e tridimensionali per grafici. La finestra di dialogo controlla anche i gradi di rotazione verticali e orizzontali per i grafici tridimensionali.

Grafico a due dimensioni
Fornisce opzioni per il controllo della stima di potenza bidimensionale versus grafici. Le impostazioni sono disabilitate per impostazione predefinita.

Confronto tra stima della potenza e valore dell'ipotesi nulla
Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per grafico a valore nullo. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e valore dell'ipotesi alternativa
Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per grafico di valore alternativo. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e la differenza tra valori ipotizzati
Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per differenza tra il grafico dei valori ipotizzati. L'impostazione è disabilitata per impostazione predefinita.

Confronto tra stima della potenza e dimensione del campione (in coppie)
Quando è abilitata, questa impostazione opzionale controlla la potenza bidimensionale per dimensione del campione. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Intervallo grafico della dimensione del campione
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per la stima di potenza bidimensionale per grafico di dimensione del campione. Il valore deve essere maggiore o uguale a 4 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima della potenza bidimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Grafico a tre dimensioni
Fornisce opzioni per il controllo della stima di potenza tridimensionale versus grafici. Per impostazione predefinita, questa impostazione è disabilitata.

Confronto tra stima della potenza e dimensione del campione
Quando selezionata, questa impostazione abilita le seguenti opzioni.

sull'asse x e la differenza tra i valori ipotizzati su asse y
L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (x -asse) e differenza tra i valori ipotizzati (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

sull'asse y e la differenza tra i valori ipotizzati su asse x
L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (asse y) e differenza tra i valori ipotizzati (x - asse). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza

Intervallo grafico della dimensione del campione (in coppie)
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per la stima di potenza tridimensionale per grafico delle dimensioni del campione. Il valore deve essere maggiore o uguale a 4 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima di potenza tridimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Confronto tra stima della potenza e valore dell'ipotesi nulla
Quando selezionata, questa impostazione abilita le seguenti opzioni.

su asse x e valore dell'ipotesi alternativa su asse y
L'impostazione opzionale controlla la potenza tridimensionale per grafico nullo (x -asse) e valore alternativo (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

su asse y e valore dell'ipotesi alternativa su asse x
L'impostazione opzionale controlla la potenza tridimensionale da parte di un grafico nullo (asse y) e valore alternativo (x -asse) .. Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Analisi di potenza del test di correlazione di Spearman a un campione

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

Il coefficiente di correlazione di grado - ordine di Spearman è una statistica non parametrica a base di rango di misura per misurare la relazione monotonica tra due variabili solitamente censurate e non normalmente distribuite. La correlazione di rango di Spearman è uguale alla correlazione di Pearson tra i valori di rango delle due variabili, e quindi anch'essa compresa tra -1 e 1. Rilevare la potenza del test di correlazione di grado Spearman è un argomento importante nell'analisi dei dati della serie temporale idrologica.

Il test utilizza il metodo asintotico di Fisher per stimare la potenza per la correlazione degli ordini di grado di uno Spearman.

1. Dai menu, scegliere:

   Analizzare > **Analisi di alimentazione** > **Correlazioni** > **Ordine di Spearman - Ordine**

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

   Per ulteriori informazioni, consultare "Analisi di alimentazione: Valori Grigi" a pagina 38.
4. Quando **Power** viene selezionata come impostazione di prova **Stima** , inserire l'apposito valore **Dimensione di esempio in coppie** . Il valore deve essere un intero intero maggiore di 3.
5. Inserire un valore che specifica il valore di ipotesi alternativo del parametro di correlazione nel campo **Parametro di correlazione Spearman** . Il valore deve essere un numero singolo compreso tra -1 e 1.

   Nota: Quando si specifica **Power**, il valore **del parametro di correlazione Spearman** non può essere -1 o 1 e non può essere uguale al **valore Null**.
6. Facoltativamente, inserire un valore che specifica il valore di ipotesi nullo del parametro di correlazione da verificare nel campo **Valore Null** . Il valore deve essere un numero singolo compreso tra -1 e 1. Il valore predefinito è 0.

   Nota: Quando si specifica **Power**, il **valore Null** non può essere -1 o 1.
7. Opzionalmente, selezionare un'opzione che determina come viene stimata la varianza asintotica per l'analisi di potenza.

   Bonett e Wright
   Stima la varianza suggerita da Bonett e Wright. Questa è l'impostazione predefinita.

   Fieller, Hartley e Pearson
   Stima la varianza suggerita da Fieller, Hartley e Pearson.

   Caruso e Cliff
   Stima la varianza suggerita da Caruso e Cliff.
8. Selezionare se il test è uno o due lati.

   Analisi non direzionale (bidirezionale)
   Quando viene selezionato, viene utilizzato un test a due lati. Questa è l'impostazione predefinita.

   Analisi direzionale (unilaterale)
   Quando selezionato, la potenza viene calcolata per un test unilaterale.
9. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.
10. Opzionalmente, clicca su **Plot** per specificare le impostazioni "Analisi di potenza di One - Sample Spearman Correlazione: Plot" a pagina 30 (output grafico, impostazioni della trama bidimensionale e impostazioni della trama tridimensionale).

    Nota: Plot è disponibile solo quando **Power** viene selezionato come presupposto di prova.
11. Opzionalmente, clicca su **Precisione** per stimare la dimensione del campione in base agli intervalli di confidenza specificando i valori dell'intervallo di confidenza di mezza larghezza. Per ulteriori informazioni, consultare "Analisi di alimentazione: Precisione" a pagina 38.

Nota: Precision è disponibile solo quando **Dimensione di esempio** viene selezionata come metodo di test supposizione **Stima** e l'analisi **Non direzionale (a due lati)** viene selezionata come **Test Direction**.

Analisi di potenza di One - Sample Spearman Correlazione: Plot

La finestra di dialogo **Plot** fornisce opzioni per il controllo dei complotti che sono in uscita per illustrare la potenza a due e tridimensionali per grafici. La finestra di dialogo controlla anche i gradi di rotazione verticali e orizzontali per i grafici tridimensionali.

Grafico a due dimensioni

Fornisce opzioni per il controllo della stima di potenza bidimensionale versus grafici. Le impostazioni sono disabilitate per impostazione predefinita.

Confronto tra stima della potenza e valore dell'ipotesi nulla

Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per grafico a valore nullo. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e valore dell'ipotesi alternativa

Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per grafico di valore alternativo. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e la differenza tra valori ipotizzati

Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per differenza tra il grafico dei valori ipotizzati. L'impostazione è disabilitata per impostazione predefinita.

Confronto tra stima della potenza e dimensione del campione (in coppie)

Quando è abilitata, questa impostazione opzionale controlla la potenza bidimensionale per dimensione del campione. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Intervallo grafico della dimensione del campione

Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore

Controlla il limite inferiore per la stima di potenza bidimensionale per grafico di dimensione del campione. Il valore deve essere maggiore o uguale a 4 e non può essere maggiore del valore **Upper bound** .

Limite superiore

Controlla il limite superiore per la stima della potenza bidimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Grafico a tre dimensioni

Fornisce opzioni per il controllo della stima di potenza tridimensionale versus grafici. Per impostazione predefinita, questa impostazione è disabilitata.

Confronto tra stima della potenza e dimensione del campione

Quando selezionata, questa impostazione abilita le seguenti opzioni.

sull'asse x e la differenza tra i valori ipotizzati su asse y

L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (x -asse) e differenza tra i valori ipotizzati (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

sull'asse y e la differenza tra i valori ipotizzati su asse x

L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (asse y) e differenza tra i valori ipotizzati (x - asse). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza

Intervallo grafico della dimensione del campione (in coppie)

Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore

Controlla il limite inferiore per la stima di potenza tridimensionale per grafico delle dimensioni del campione. Il valore deve essere maggiore o uguale a 4 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima di potenza tridimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Confronto tra stima della potenza e valore dell'ipotesi nulla
Quando selezionata, questa impostazione abilita le seguenti opzioni.

su asse x e valore dell'ipotesi alternativa su asse y
L'impostazione opzionale controlla la potenza tridimensionale per grafico nullo (x -asse) e valore alternativo (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

su asse y e valore dell'ipotesi alternativa su asse x
L'impostazione opzionale controlla la potenza tridimensionale da parte di un grafico nullo (asse y) e valore alternativo (x -asse) .. Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Analisi di alimentazione del Partial Pearson Correlazione Test

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

La correlazione parziale può essere spiegata come l'associazione tra due variabili casuali dopo l'eliminazione dell'effetto di un'altra o di diverse altre variabili. Si tratta di una misurazione utile in presenza di confondazione. Analogamente al coefficiente di correlazione di Pearson, anche il coefficiente di correlazione parziale è una quantità adimensionale che varia tra -1 e 1.

Il test utilizza il metodo asintotico di Fisher per stimare la potenza per la correlazione di Pearson a campione.

1. Dai menu, scegliere:

   Analizzare > **Analisi Alimentazione** > **Correlazioni** > **Parziale**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

   Per ulteriori informazioni, consultare “Analisi di alimentazione: Valori Grigi” a pagina 38.

4. Quando **Power** viene selezionata come impostazione di prova **Stima** , inserire l'apposito valore **Dimensione di esempio** . Il valore deve essere un intero intero maggiore di 1.
5. Inserire un valore che specifica il **Numero delle variabili ipotiate per essere partigiate**. Il valore deve essere un intero intero maggiore o uguale a 0.
6. Inserire un valore che specifica il valore di ipotesi alternativo del **parametro di correlazione parziale**. Il valore deve essere un singolo numero compreso tra -1 e 1.

   Nota: Quando si specifica **Power**, il valore **del parametro Correlazione parziale** non può essere -1 o 1 e non può essere uguale al **valore Null**.
7. Facoltativamente, inserire un valore che specifica il valore di ipotesi nullo del parametro di correlazione parziale da verificare nel campo **Valore Null** . Il valore deve essere un singolo numero compreso tra -1 e 1. Il valore predefinito è 0.

   Nota: Quando si specifica **Power**, il **valore Null** non può essere -1 o 1.
8. Selezionare se il test è uno o due lati.

   Analisi non direzionale (bidirezionale)
   Quando viene selezionato, viene utilizzato un test a due lati. Questa è l'impostazione predefinita.

   Analisi direzionale (unilaterale)
   Quando selezionato, la potenza viene calcolata per un test unilaterale.
9. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.
10. Opzionalmente, clicca su **Plot** per specificare le impostazioni “Power Analysis di Parziale Pearson Correlazione: Plot” a pagina 33 (output grafico, impostazioni della trama bidimensionale e impostazioni della trama tridimensionale).

    Nota: Plot è disponibile solo quando **Power** viene selezionato come presupposto di prova.
11. Opzionalmente, clicca su **Precisione** per stimare la dimensione del campione in base agli intervalli di confidenza specificando i valori dell'intervallo di confidenza di mezza larghezza. Per ulteriori informazioni, consultare “Analisi di alimentazione: Precisione” a pagina 38.

Nota: Precision è disponibile solo quando **Dimensione di esempio** viene selezionata come metodo di test supposizione **Stima** e l'analisi **Non direzionale (a due lati)** viene selezionata come **Test Direction**.

Power Analysis di Parziale Pearson Correlazione: Plot

La finestra di dialogo **Plot** fornisce opzioni per il controllo dei complotti che sono in uscita per illustrare la potenza a due e tridimensionali per grafici. La finestra di dialogo controlla anche i gradi di rotazione verticali e orizzontali per i grafici tridimensionali.

Grafico a due dimensioni
Fornisce opzioni per il controllo della stima di potenza bidimensionale versus grafici. Le impostazioni sono disabilitate per impostazione predefinita.

Confronto tra stima della potenza e valore dell'ipotesi nulla
Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per grafico a valore nullo. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e valore dell'ipotesi alternativa
Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per grafico di valore alternativo. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e numero di variabili rese parziali
Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per numero di variabili di partimurate. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e la differenza tra valori ipotizzati
Quando abilitata, questa impostazione opzionale controlla la potenza bidimensionale per differenza tra il grafico dei valori ipotizzati. L'impostazione è disabilitata per impostazione predefinita.

Confronto tra stima della potenza e dimensione del campione
Quando è abilitata, questa impostazione opzionale controlla la potenza bidimensionale per dimensione del campione. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Intervallo grafico della dimensione del campione
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per la stima di potenza bidimensionale per grafico di dimensione del campione. Il valore deve essere maggiore o uguale a 4 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima della potenza bidimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Grafico a tre dimensioni
Fornisce opzioni per il controllo della stima di potenza tridimensionale versus grafici. Per impostazione predefinita, questa impostazione è disabilitata.

Confronto tra stima della potenza e dimensione del campione
Quando selezionata, questa impostazione abilita le seguenti opzioni.

sull'asse x e la differenza tra i valori ipotizzati su asse y
L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (x -asse) e differenza tra i valori ipotizzati (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

sull'asse y e la differenza tra i valori ipotizzati su asse x
L'impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (asse y) e differenza tra i valori ipotizzati (x - asse). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza

Intervallo grafico della dimensione del campione (in coppie)
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per la stima di potenza tridimensionale per grafico delle dimensioni del campione. Il valore deve essere maggiore o uguale a 4 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima di potenza tridimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Confronto tra stima della potenza e valore dell'ipotesi nulla
Quando selezionata, questa impostazione abilita le seguenti opzioni.

su asse x e valore dell'ipotesi alternativa su asse y
L'impostazione opzionale controlla la potenza tridimensionale per grafico nullo (x -asse) e valore alternativo (asse y). Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

su asse y e valore dell'ipotesi alternativa su asse x
L'impostazione opzionale controlla la potenza tridimensionale da parte di un grafico nullo (asse y) e valore alternativo (x -asse) .. Per impostazione predefinita, il grafico è soppresso. Quando specificato, il grafico visualizza.

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Regressione

Le seguenti funzioni statistiche sono incluse in IBM SPSS Statistica Base Edition.

Analisi di potenza di Univariate Linear Regression Test

Questa funzione richiede IBM SPSS Statistica Base Edition.

L'analisi di potenza svolge un ruolo cardine in un piano di studio, design e conduzione. Il calcolo del potere è di solito prima che siano stati raccolti dati di campione, salvo possibilmente da un piccolo studio pilota. La stima precisa del potere può dire agli investigatori quanto sia probabile che una differenza statisticamente significativa venga rilevata in base a una dimensione di campione finito sotto una vera ipotesi alternativa. Se il potere è troppo basso, ci sono poche possibilità di rilevare una differenza significativa, e i risultati non significativi sono probabili anche se davvero le differenze reali esistono.

La regressione lineare univariata è un approccio statistico di base e standard in cui i ricercatori utilizzano i valori di diverse variabili per spiegare o prevedere valori di un risultato su scala.

Il test Power Analysis of Univariate Linear Regression stima la potenza del tipo III *F*- test in univariati modelli di regressione lineare multipli. Con la dimensione dell'effetto rappresentata da più (parziali) correlazioni, vengono forniti approcci per i predittori fissi e casuali. Per i predittori fissi, la stima dell'alimentazione si basa sulla distribuzione *F*non centrale. Per i predittori casuali si ipotizza che la variabile target e i predittori seguano congiuntamente una distribuzione normale multivariata. In tal caso la stima dell'alimentazione si basa sulla distribuzione del coefficiente di correlazione multiplo di campione.

1. Dai menu, scegliere:

   Analizzare > **Analisi Alimentazione** > **Regressione** > **Univariate Linear**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare un'impostazione di test **Stima** (**Dimensione di esempio** o **Power**).
3. Quando **Dimensione di esempio** viene selezionata, inserire un valore di stima della dimensione di esempio di **Single power value** (il valore deve essere un valore unico compreso tra 0 e 1) oppure selezionare **Valori di alimentazione Griglia** e quindi fare clic su **Griglia** per visualizzare le dimensioni del campione proiettate per una gamma di valori di alimentazione specifici.

   Per ulteriori informazioni, consultare “Analisi di alimentazione: Valori Grigi” a pagina 38.
4. Quando **Potenza stimata** viene selezionata, inserire un'apposita **dimensione di esempio** per il valore di stima dell'alimentazione. Il valore deve essere un intero intero maggiore o uguale al numero totale di predittori modello + 2 (quando **Include il termine di intercettazione nel modello** è abilitato). Altrimenti il valore deve essere un intero intero maggiore o uguale al numero totale dei predittori modello +1.

5. Specificare il valore del coefficiente di correlazione parziale multiplo nel campo **Correlazione Multipla multipla** . Il valore deve essere un singolo valore compreso tra -1 e 1.

   Nota: Quando viene specificato un valore **Power** , il valore **Correlazione Multipla multipla** non può essere 0.

   Le seguenti impostazioni sono abilitate quando si seleziona **Correlazione Multipla multipla** :

   Numero totale di predittori nel modello
   Specificare il numero dei predittori totali, o i predittori nel modello completo (non incluso l'intercettazione, se applicabile). Il valore deve essere un intero intero maggiore o uguale a 1.

   Numero di predittori di test
   Specificare il numero dei predittori di prova, o i predittori nel modello nidificato (non incluso l'intercettazione, se applicabile). Il valore deve essere maggiore o uguale a 1, ma non superiore al valore **Numero totale di predittori nel modello** .

6. Specificare **R - quadrato valori per** coefficienti di correlazione multipli per entrambi **Modello completo** e **Modello Nest**. I valori devono essere valori unici compresi tra 0 e 1.

   Nota: Quando viene specificato un valore **Power** , il valore **Full model** deve essere maggiore del valore **Nested model** .

   Le seguenti impostazioni sono abilitate quando vengono selezionati **R - quadrato valori per** :

   Numero totale di predittori- Modello completo
   Specificare il numero di predittori totali per il modello completo (non incluso l'intercettazione, se applicabile). Il valore deve essere un intero intero maggiore o uguale a 1.

   Numero totale di predittori- Modello nidificato
   Specificare il numero di predittori totali per il modello nidificato (non incluso l'intercettazione, se applicabile). Il valore deve essere maggiore o uguale a 1, ma inferiore al valore **Total number of predictors - Full model** .

7. Facoltativamente, specificare il livello di significatività della frequenza di errore di tipo I per il test nel campo **Livello di Significanza** . Il valore deve essere un solo doppio valore compreso tra 0 e 1. Il valore predefinito è 0.05.
8. È possibile selezionare facoltativamente la impostazione **Includi il termine di intercettazione nell'impostazione del modello** . L'impostazione è abilitata per impostazione predefinita. Quando non viene selezionato, il termine di intercettazione è escluso dall'analisi di alimentazione.
9. È possibile opzionalmente selezionare se i predittori del modello sono **Fixed** o **Random**. **Fisso** è l'impostazione predefinita.
10. È possibile opzionalmente fare clic su **Plot** per specificare le impostazioni “Analisi di potenza di Regressione Lineare Univariata: Trama” a pagina 36 (output grafico, impostazioni della trama bidimensionale e impostazioni della trama tridimensionale).

    Nota: Plot è disponibile solo quando **Power** viene selezionato come presupposto di prova.

Analisi di potenza di Regressione Lineare Univariata: Trama

È possibile controllare i complotti che sono in uscita per illustrare la potenza a due e tridimensionali mediante grafici. È inoltre possibile controllare la visualizzazione dei suggerimenti degli utensili e dei gradi di rotazione verticali / orizzontali per i grafici tridimensionali.

Grafico a due dimensioni
Fornisce opzioni per il controllo della stima di potenza bidimensionale versus grafici. Le impostazioni sono disabilitate per impostazione predefinita.

Confronto tra stima della potenza e la correlazione parziale multipla
Quando consente, questa impostazione opzionale controlla la potenza bidimensionale tramite più grafico a coefficiente di correlazione parziale. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Confronto tra stima della potenza e dimensione del campione
Quando è abilitata, questa impostazione opzionale controlla la potenza bidimensionale per dimensione del campione. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Intervallo grafico della dimensione del campione
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per la stima di potenza bidimensionale per grafico di dimensione del campione. Il valore deve essere maggiore o uguale a 4 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima della potenza bidimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Grafico a tre dimensioni
Fornisce opzioni per il controllo della stima di potenza tridimensionale versus grafici, le impostazioni di rotazione verticale / orizzontale e la gamma di tracciamento di dimensioni del campione specificato dall'utente. Per impostazione predefinita, questa impostazione è disabilitata.

Confronto tra stima della potenza e dimensione del campione
Quando abilitata, questa impostazione opzionale controlla la potenza tridimensionale per grafici di dimensione del campione. L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

su asse x e la correlazione parziale multipla su asse y
Questa impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (x - asse) e grafico a coefficiente di correlazione parziale multiplo (asse y). L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

su asse y e la correlazione parziale multipla su asse x
Questa impostazione opzionale controlla la potenza tridimensionale per dimensione del campione (asse y) e grafico a coefficiente di correlazione parziale multiplo (x - asse). L'impostazione è disabilitata per impostazione predefinita. Quando selezionata, questa impostazione visualizza il grafico.

Intervallo grafico della dimensione del campione
Quando selezionate, sono disponibili le opzioni di collegamento più basse e superiori. Quando non vengono specificati valori interi per i campi **Lower bound** o **Upper bound** , viene utilizzata la gamma di trama predefinita.

Limite inferiore
Controlla il limite inferiore per la stima di potenza bidimensionale per grafico di dimensione del campione. Il valore deve essere maggiore o uguale a 4 e non può essere maggiore del valore **Upper bound** .

Limite superiore
Controlla il limite superiore per la stima della potenza bidimensionale per dimensione del campione. Il valore deve essere maggiore del valore **Bassa bound** e non può essere maggiore di 5000.

Rotazione verticale
L'impostazione opzionale imposta i gradi di rotazione verticale (in senso orario da sinistra) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare verticalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 10.

Rotazione orizzontale
L'impostazione opzionale imposta i gradi di rotazione orizzontali (in senso orario dalla parte anteriore) per il grafico tridimensionale. È possibile utilizzare il mouse per ruotare orizzontalmente il grafico. L'impostazione prende effetto quando viene richiesta la trama tridimensionale. Il valore deve essere un singolo numero intero inferiore o uguale a 359. Il valore predefinito è 325.

Analisi di alimentazione: Precisione

La finestra di dialogo **Precisione** è disponibile per ogni procedura di Power Analysis (ad eccezione di Univariate Linear Regression), e fornisce opzioni per stimare la dimensione del campione in base all'intervallo di confidenza specificato per metà larghezza. La finestra di dialogo è disponibile quando **Valore di potenza singola** o **Valori di alimentazione Griglia** vengono specificati per la stima **Dimensione di esempio** e quando l'impostazione **Verifica Direzione** è impostata su **Nondirezionale (analisi a due lati)**.

Nota: L'impostazione **Verifica Direzione** non si applica a One-Way ANOVA.

quando le opzioni **Dimensione del campione preventivo** e **Valori di alimentazione della griglia** sono selezionate (clicca sul controllo **Griglia** per visualizzare la finestra di dialogo).

Tipo(i) intervallo di confidenza
Selezionare i tipi di intervalli di confidenza appropriati e se viene applicata la correzione di continuità.

Nota: Interval Interval Type (s) sono disponibili solo per le procedure di proporzione di Power Analysis (One - Sample Proporzione, Related - Samples Binomial Test, and Independent - Samples Binomial Test).

Specificare la metà - larghezza dell'intervallo di confidenza
Stima la dimensione del campione in base al valore di mezza larghezza di confidenza. Inserire un valore nella gamma 0 - 1. Per il test binomiale a campione singolo, il valore deve essere compreso nell'intervallo 0 - 0.5.

Nota: I valori duplicati vengono ignorati.

- Clicca su **Aggiungi** per aggiungere il valore di mezza larghezza specificato all'elenco.
- Evidenziare un valore di mezza larghezza esistente e fare clic su **Modifica** per aggiornare il valore.
- Evidenziare un valore di mezza larghezza esistente e fare clic su **Rimuovi** per rimuovere il valore dall'elenco.

Analisi di alimentazione: Valori Grigi

La finestra di dialogo **Valori griglia** fornisce opzioni per specificare un intervallo di valori **POWER** allo scopo di visualizzare le dimensioni del campione previste in un formato griglia per ogni valore di intervallo **POWER** specificato.

La finestra di dialogo **Valori di griglia** è disponibile per ogni procedura di Power Analysis quando le opzioni **Dimensione del campione preventivo** e **Valori di alimentazione della griglia** sono selezionate (clicca sul controllo **Griglia** per visualizzare la finestra di dialogo).

Specifica potenza singola
Quando viene selezionato, è necessario almeno un valore per eseguire l'analisi. Sono consentiti più valori e ogni valore deve appartenere a `[0, 1]`. È possibile specificare più valori con ogni valore separato da uno spazio vuoto (o spazi vuoti). Utilizzare i controlli **Aggiungi**, **Modifica**e **Rimuovi** per lavorare con i valori nella lista dei valori di potenza.

Tutti i valori devono essere univeri (non sono consentiti valori duplicati).

Specifica intervallo potenza
Quando selezionato, è possibile specificare una gamma di valori di potenza da un valore **Start** (`value1`) ad un valore **Fine** (`value2`) con l'incremento di **By** (`value3`). Se specificato, è consentito solo una serie valida di [`value1 TO value2 BY value3`]. Deve soddisfare tale `0 ≤ value1 ≤ value2 ≤ 1`. Nei casi in cui `value1 = value2`, equivale a specificare un singolo `value1`, indipendentemente da `value3`.

Nota: Le opzioni **Specifica alimentazione singola** e **Specifica potenza** sono indipendenti; è possibile selezionare un'unica opzione o entrambe le opzioni.

Meta - analisi

Meta analysis è l'analisi dei dati ottenuti da una raccolta di studi che rispondono a domande di ricerca analoghe. Questi studi sono noti come studi primari. Meta analisi utilizza metodi statistici per produrre una stima complessiva di un effetto, esplorare l'eterogeneità di studio e indagare l'impatto della pubblicazione bias o, più in generale, effetti di piccole dimensioni sui risultati finali.

IBM SPSS Statistica supporta dimensioni di effetto standard e dimensioni di effetto generiche (pre - calcolate) per entrambi i dati binari (come ad esempio il odds ratio - ratio) e per i dati continui (come ad esempio l'g di Hedges). Le informazioni di meta analisi (come ad esempio le dimensioni degli effetti specifici dello studio e i relativi errori standard corrispondenti e il metodo e il metodo di analisi meta-analisi) vengono specificate durante la fase di analisi della meta analisi. Le informazioni vengono utilizzate automaticamente da tutte le mete successive.

Effetti casuali, effetto comune e modelli di meta analisi effetti fissi sono supportati. A seconda del modello di meta analisi prescelto, i vari metodi di stima (ad esempio inverse - variance e Mantel-Haenszel) sono disponibili per l'effetto comune e i modelli di effetti fissi. Diversi stimatori diversi sono disponibili anche per il parametro di varianza tra studio per il modello di effetti casuali.

Meta - Analisi Continuo

La Meta - Analysis Continuous esegue la meta - analisi con esiti continui su dati grezzi che vengono forniti nel dataset attivo per la stima della dimensione dell'effetto.

Esempio

Diversi studi di ricerca sono stati condotti nella storia per indagare su una medicina faddish ma discutibile per aiutare a trattare il diabete di tipo II. La medicina orale è stata pretesa di poter ridurre il livello di glucosio nel sangue dopo i pasti. I dati sono stati raccolti da diversi siti di ricerca dal 1979 al 1986.

Un investigatore principale vorrebbe trarre inferenza statistica sull'effetto della medicina orale. A causa del fatto che i dati sono stati generati da diversi studi, ha proposto l'idea di sintetizzare i risultati attraverso gli studi per raggiungere una comprensione complessiva dell'effetto e individuare quelle sottostanti fonti di variazione degli esiti.

Statistiche

Livello di confidenza, metodo iterativo, step-halving, tolleranza di convergenza, media campionaria, varianza campionaria, deviazione standard, dimensione stimata dell'effetto, *d* di Cohen, *g* di Hedges, delta di Glass, differenza media, analisi cumulativa, metodo di stima, trim-and-fill, test basato sulla regressione, modello a effetti casuali, modello a effetti fissi, stimatore a massima verosimiglianza ristretto, stimatore empirico di Bayes, stimatore di Hedges, stimatore di Hunter-Schmidt, DerSimonian-Laird stimatore, stimatore di Sidik-Jonkman, aggiustamento dell'errore standard di Knapp-Hartung, aggiustamento dell'errore standard di Knapp-Hartung troncato, coefficienti, test di EGGER basato sulla regressione, intercetta, modello moltiplicativo, parametro di dispersione moltiplicativo, stimatore quadratico, test di omogeneità, misure di eterogeneità, intervallo di previsione, errore standard stimato, *p-value* stimato, dimensione dell'effetto complessivo cumulativo, peso stimato dello studio.

Ottenimento di un'analisi Continuous Meta - Analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dati raw ...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Nella sezione **Gruppo di trattamento** , selezionare una variabile **Dimensione studio** per rappresentare la dimensione del campione per il gruppo di trattamento. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
3. Selezionare una variabile **Mean** per rappresentare i mezzi di esempio per il gruppo di trattamento. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
4. Selezionare **deviazione standard** per specificare la deviazione standard del campione oppure **Varianza** per specificare la varianza di esempio e quindi selezionare una variabile per rappresentare la deviazione standard/varper il gruppo di trattamento.
5. Nella sezione **Gruppo di controllo** , selezionare una variabile **Dimensione studio** per rappresentare la dimensione del campione per il gruppo di controllo. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
6. Selezionare una variabile **Mean** per rappresentare i mezzi di esempio per il gruppo di controllo. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
7. Selezionare **deviazione standard** per specificare la deviazione standard del campione oppure **Varianza** per specificare la varianza di esempio e quindi selezionare una variabile per rappresentare la deviazione standard/varper il gruppo di controllo.
8. Facoltativamente, selezionare le variabili **Studio ID** e / o **Label di studio** . La variabile **Studio ID** selezionata non può essere la stessa della variabile selezionata **Studio Label** .
9. Facoltativamente, selezionare un'impostazione **Effect Size** .

   D di Cohen
   L'impostazione predefinita stima il *d*di Cohen. Quando viene selezionato **Errore standard corretto** , l'impostazione stima il *d* di Cohen e la sua varianza utilizzando una formula alternativa divisa da $2(N_{treatment} + N_{control} - 2)$.

   G di Hedges
   Stima gli Hedges ' *g*. Quando viene selezionato **Errore standard corretto** , l'impostazione stima gli Hedges ' *g* e la sua varianza utilizzando una formula alternativa divisa da $2(N_{treatment} + N_{control} - 3.94)$.

   Delta di Glass
   Stima il Delta del vetro basato sul gruppo di controllo. Quando **Standardized basato sul gruppo di trattamento** viene selezionato, il Delta del vetro è standardizzato in base alla deviazione standard del gruppo di trattamento.

   Differenza media non standardizzata
   Stima la differenza media ipotizzando le due deviazioni standard di popolazione sono uguali. Quando **le varianze di gruppo Unequal** viene selezionata la differenza media è stimata ipotizzando che le due deviazioni standard della popolazione non siano uguali.
10. Facoltativamente, selezionare un'impostazione **Modello** . Quando le impostazioni **Trim - e - Fill** sono abilitate, l'impostazione controlla anche il modello utilizzato dal pooling nell'analisi del trim - e - fill. Quando le impostazioni **Bias** sono abilitate, l'impostazione controlla anche il modello utilizzato dal test basato sulla regressione.

    Effetti casuali
    L'impostazione predefinita costruisce il modello di effetti casuali.

    Effetti fissi
    Costruisce il modello a effetti fissi.
11. Facoltativamente, è possibile:
    - Clicca su **Criteri ...** per specificare i criteri generali.
    - Fare clic su **Analisi** per specificare il sottogruppo e l'analisi cumulativa.
    - Clicca su **Inferenza** per specificare i metodi di stima.
    - Clicca su **Contrast** per controllare il test di contrasto.
    - Clicca su **Bias** per accedere alla bias di pubblicazione conducendo il test di regressione di EGGER.
    - Clicca su **Trim - e - Fill** per implementare l'analisi di trim e - fill di pubblicazione bias.

- Clicca su **Stampa** per controllare le uscite della tabella.
- Clicca su **Salva** per salvare le statistiche stimate al dataset attivo.
- Fare clic su **Plot** per specificare quali grafici includere nell'output.

12. Fare clic su **OK**.

Meta - Analisi Continua: Criteri

La finestra di dialogo **Criteri** fornisce le impostazioni per specificare i criteri di meta-analisi con esiti continui su dati greli forniti nel dataset attivo per la stima della dimensione dell'effetto.

Intervallo di confidenza
L'impostazione opzionale specifica il livello di confidenza. Il valore deve essere un valore numerico compreso tra 0 e 100. L'impostazione predefinita è 95.

Ambito dati mancanti
Le impostazioni opzionali controllano come la procedura gestisce i dati mancanti.

Escludi casi analisi per analisi
L'impostazione predefinita include tutti i casi con dati sufficienti sulle variabili utilizzate in ogni particolare analisi.

Escludi casi listwise
Include tutti i casi con dati sufficienti su tutte le variabili utilizzate in tutte le analisi specificate dalla procedura.

Valori mancanti definiti dall'utente
Le impostazioni opzionali controllano come vengono trattati i valori mancanti dell'utente.

Escludi
L'impostazione predefinita tratta i valori mancanti dell'utente come validi.

Includi
Ignora le designazioni di valore mancanti dell'utente e tratta i valori mancanti dell'utente come validi.

Iterazioni

Numero massimo di iterazioni
L'impostazione opzionale specifica il numero massimo di iterazioni nei metodi iterativi. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore deve essere un intero intero positivo. Il valore predefinito è 100. Un valore di 0 significa che non vengono eseguite iterazioni.

Numero massimo di dimezzamenti
L'impostazione opzionale specifica il passo massimo - dimezzamento nei metodi iterativi. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore deve essere un intero intero positivo. Il valore predefinito è 5. Un valore di 0 significa che il dimezzamento non viene applicato.

Convergenza
L'impostazione opzionale specifica la tolleranza di convergenza. Il valore deve essere un valore positivo unico. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore predefinito è 1E-6.

Definizione Dei Criteri Di Meta - Analisi Continua

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dati raw ...**

2. Nella finestra di dialogo **Meta - Analisi Continua** , fare clic su **Criteri**.
3. Selezionare e definire le impostazioni dei criteri appropriati.
4. Fare clic su **Continua**.

Meta - Analisi Continuo: Analisi

La finestra di dialogo **Analisi** fornisce le impostazioni per la specifica del sottogruppo e l'analisi cumulativa per la meta - analisi con esiti continui su dati greli forniti nel dataset attivo per la stima della dimensione dell'effetto.

Analisi del sottogruppo
Selezionare una variabile che evoca l'analisi del sottogruppo. La variabile non può essere uguale alla variabile definita per **Analisi cumulativa**.

Analisi cumulativa
Seleziona una variabile che evoca l'analisi cumulativa e su cui viene condotta la meta - analisi cumulativa. La variabile non può essere uguale alla variabile definita per **Analisi Sottogruppo**. Quando **Ascending** viene selezionato, l'analisi cumulativa si basa sulla variabile specificata in ordine crescente. Quando **Descending** viene selezionato, l'analisi cumulativa si basa sulla variabile specificata in ordine decrescente.

Statistiche cumulative
Fornisce opzioni per il salvataggio delle dimensioni complessive complessive stimate. Le impostazioni sono disponibili solo quando viene selezionata una variabile **Analisi cumulativa** .

Dimensione effetto cumulativa
Salva le dimensioni complessive complessive stimate.

Errore standard
Salva l'errore standard stimato della dimensione complessiva dell'effetto cumulativo.

Limite inferiore intervallo di confidenza
Salva l'intervallo di confidenza stimato inferiore della dimensione complessiva dell'effetto cumulativo.

Limite superiore intervallo di confidenza
Salva l'intervallo di confidenza stimato per il limite superiore della dimensione complessiva dell'effetto cumulativo.

Valore P
Salva il valore stimato *p*della dimensione complessiva dell'effetto cumulativo.

Destinazione
Fornisce le opzioni per specificare un dataset di salvataggio o un file di dati. Quando **Dataset** viene selezionato, è possibile specificare un nuovo nome dataset (è possibile conservare il nome del dataset predefinito). Quando si seleziona **File di dati** , fare clic su **Sfoglia ...** per selezionare un nome file di salvataggio e la posizione.

Definizione Impostazioni di analisi della meta - analisi continua

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dati raw ...**
2. Nella finestra di dialogo **Meta - Analysis Continuous** , fare clic su **Analisi**.
3. Selezionare e definire le impostazioni di analisi appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continua: Inferenza

La finestra di dialogo **Inferenza** fornisce le impostazioni per specificare i metodi di stima per la meta - analisi con esiti continui su dati greli forniti nel dataset attivo per la stima della dimensione dell'effetto.

Nota: La finestra di dialogo **Inferenza** è disponibile solo quando viene selezionato un modello **Random - effects** .

Stimatore
Fornisce le impostazioni per la specifica dello stimatore.

Note:

- Quando vengono specificate le impostazioni di Trim e Fill, controlla anche l'estimatore che viene utilizzato dal pooling nell'analisi del trim e del riempimento.
- Quando le impostazioni di Bias vengono specificate, controlla anche l'estimatore che viene utilizzato dal test basato sulla regressione.

Verosimiglianza massima limitata (REML)
L'impostazione predefinita applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza limitato.

Verosimiglianza massima (ML)
Applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza.

Bayes empirico
Applica il metodo iterativo e calcola lo stimatore di Bayes empirico.

Hedges
Applica il metodo non iterativo e calcola lo stimatore Hedges.

Hunter-Schmidt
Applica il metodo non iterativo e calcola lo stimatore Hunter - Schmidt.

DerSimonian-Laird
Applica il metodo non iterativo e calcola lo stimatore DerSimonian-Laird.

Sidik-Jonkman
Applica il metodo non iterativo e calcola lo stimatore Sidik-Jonkman.

Adattamento dell'errore standard
Fornisce le impostazioni che controllano se applicare la regolazione degli errori standard Knapp - Hartung.

Note:

- Quando vengono specificate le impostazioni di Trim e Fill, controlla anche la regolazione degli errori standard utilizzata dal pooling nell'analisi del trim e del riempimento.
- Quando le impostazioni di Bias vengono specificate, controlla anche la regolazione degli errori standard utilizzata dal test basato sulla regressione.

Nessun adattamento
L'impostazione predefinita non applica la regolazione.

Applica adattamento Knapp-Hartung
Applica il metodo di regolazione Knapp - Hartung.

Applica adattamento Knapp-Hartung troncato
Applica il metodo di regolazione Knapp - Hartung e tronca il valore se inferiore a 1 quando si stima la matrice di varianza - covarianza.

Definizione Impostazioni di inferenza continua di analisi meta

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dati raw ...**
2. Nella finestra di dialogo **Meta - Analysis Continuous** , clicca su **Inferenza**.
3. Selezionare e definire le impostazioni di inferenza appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continua: Contrasto

La finestra di dialogo **Contrast** fornisce le impostazioni per il controllo del test di contrasto per la meta - analisi con esiti continui su dati greli forniti nel dataset attivo per la stima della dimensione dell'effetto.

Variabili
L'elenco visualizza tutte le variabili di dataset disponibili. Selezionare le variabili dall'elenco e spostarle nella lista **Contrast (s)** .

Contrasto (s)
L'elenco identifica i coefficienti che vengono memorizzati come variabili nel dataset attivo. Sono consentite più variabili. Le variabili di stringa non sono supportate.

Valori del coefficiente immesso dall'utente
Fornisce le impostazioni per la specifica dei coefficienti di un contrasto fornito dall'utente. Sono ammessi solo valori numerici. Per formulare un test di contrasto valido, il numero di valori specificati deve corrispondere a quello degli studi validi.

Definizione Meta - Analisi Impostazioni di contrasto continui

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dati raw ...**
2. Nella finestra di dialogo **Meta - Analysis Continuous** , fare clic su **Contrast**.
3. Selezionare e definire le impostazioni di contrasto appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continuo: Bias

La finestra di dialogo **Bias** fornisce le impostazioni per l'abilitazione della bias di pubblicazione conducendo il test di regressione di Eggers per la meta - analisi con esiti continui su dati raw forniti nel dataset attivo per la stima della dimensione dell'effetto.

Test basato sulla regressione di Egger
La selezione di questa impostazione abilita la bias di pubblicazione conducendo il test di regressione di Eggers.

Variabili
L'elenco fornisce tutte le variabili di dataset disponibili.

Covariate
Le variabili selezionate dall'elenco **Variabili** sono trattate come covariate. Sono consentite più covariate.

Fattori
Le variabili selezionate dall'elenco **Variabili** sono trattate come fattori. Sono consentiti più fattori.

Includi intercettazione nella regressione
Controlla il termine di intercettazione nel test di regressione.

Includi parametro di dispersione nel modello a effetti fissi
Controlla l'impostazione del modello moltiplicativo e introduce il parametro di dispersione moltiplicativo all'analisi. L'impostazione è disponibile solo quando viene selezionato un modello a effetti fissi.

Stima statistiche in base a distribuzione t
Controlla la distribuzione utilizzata nei test di regressione. L'impostazione è abilitata per impostazione predefinita, che stima le statistiche in base alla distribuzione *t*. Quando l'impostazione non viene selezionata, le statistiche vengono stimate in base alla distribuzione normale.

Definizione Impostazioni di bias a meta - analisi continua

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dati raw ...**
2. Nella finestra di dialogo **Meta - Analysis Continuous** , fare clic su **Bias**.
3. Selezionare e definire le impostazioni di bias appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continuo: Trim - and - Fill

La finestra di dialogo **Trim - and - Fill** fornisce le impostazioni per implementare l'analisi trim - and - fill di pubblicazione bias per la meta - analisi con esiti continui su dati raw forniti nel dataset attivo per la stima della dimensione dell'effetto.

Stima numero di studi mancanti
Controlla l'analisi del trim e - fill di pubblicazione bias. La selezione di questa impostazione abilita le altre impostazioni di dialogo.

Lato per assegnare gli studi
Fornisce opzioni per specificare il lato della trama del funnel su cui sono imputati gli studi mancanti.

Determinato dalla pendenza del test di Egger
L'impostazione predefinita determina il lato in base alla pendenza stimata del test di Egger.

A sinistra
Imputa il lato sinistro della trama del funnel.

Destra
Imputa il lato destro della trama del funnel.

Metodo
Specifica il metodo per stimare il numero di studi mancanti.

Lineare
L'impostazione predefinita calcola lo stimatore lineare.

Esegui
Calcola lo stimatore di esecuzione.

Quadratico
Calcola lo stimatore quadratico.

Processo di iterazione
Fornisce le impostazioni per specificare l'indicatore di iterazione e la regolazione degli errori standard.

Modello a effetti fissi
Quando viene selezionato, viene impiegato un modello a effetti fissi e la stima dell'iterazione e le opzioni di regolazione degli errori standard non sono disponibili.

Modello a effetti casuali
Quando viene selezionato, è impiegato un modello di effetti casuali e sono disponibili le seguenti impostazioni.

Stimatore
Fornisce le impostazioni per la specifica dello stimatore di iterazione

Verosimiglianza massima limitata (REML)
L'impostazione predefinita applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza limitato.

Verosimiglianza massima (ML)
Applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza.

Bayes empirico
Applica il metodo iterativo e calcola lo stimatore di Bayes empirico.

Hedges
Applica il metodo non iterativo e calcola lo stimatore Hedges.

Hunter-Schmidt
Applica il metodo non iterativo e calcola lo stimatore Hunter - Schmidt.

DerSimonian-Laird
Applica il metodo non iterativo e calcola lo stimatore DerSimonian-Laird.

Sidik-Jonkman
Applica il metodo non iterativo e calcola lo stimatore Sidik-Jonkman.

Adattamento dell'errore standard
Fornisce le impostazioni per il controllo se applicare l'adattamento di errore standard Knapp - Hartung alle iterazioni dell'algoritmo di trim e riempimento.

Nessun adattamento
L'impostazione predefinita non applica la regolazione.

Applica adattamento Knapp-Hartung
Applica il metodo di regolazione Knapp - Hartung.

Applica adattamento Knapp-Hartung troncato
Applica il metodo di regolazione Knapp - Hartung e tronca il valore se inferiore a 1 quando si stima la matrice di varianza - covarianza.

Definizione Meta - Analisi Continue impostazioni di trim - e - fill

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dati raw ...**
2. Nella finestra di dialogo **Meta - Analisi Continuous** , fare clic su **Trim - e - Fill**.
3. Selezionare e definire le impostazioni di allestimento e di riempimento appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continuo: Stampa

La finestra di dialogo **Stampa** fornisce le impostazioni per il controllo degli output della tabella per la meta - analisi con esiti continui su dati greli forniti nel dataset attivo per la stima della dimensione dell'effetto.

Omogeneità/eterogeneità
Fornisce le impostazioni per il controllo dei test di omogeneità e eterogeneità.

Test di omogeneità
Quando viene selezionato, nella produzione viene fornito il corrispondente test di omogeneità.

Misure di eterogeneità
Quando selezionate, le misure di eterogeneità sono fornite nell'output.

Dimensioni effetto
Fornisce le seguenti impostazioni di dimensione dell'effetto.

Studi individuali
Controlla la visualizzazione dei singoli studi. Quando viene selezionato, viene fornita la corrispondente emissione.

Dimensioni effetto cumulativo
Controlla la visualizzazione dell'analisi cumulativa. Quando viene selezionato, viene fornita la corrispondente emissione. L'impostazione è disponibile solo quando una variabile **Analisi cumulativa** viene selezionata sulla finestra di dialogo **Analisi** .

Intervallo di previsione nel modello a effetti casuali
Controlla la visualizzazione dell'intervallo di previsione. Quando viene selezionato, viene fornita la corrispondente emissione. L'impostazione è disponibile solo quando viene specificato un modello a effetti casuali.

Definizione Meta - Analisi Impostazioni di stampa continua

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dati raw ...**
2. Nella finestra di dialogo **Meta - Analisi Continua** , fare clic su **Stampa**.
3. Selezionare e definire le impostazioni di stampa appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continuo: Salva

La finestra di dialogo **Salva** fornisce le impostazioni per il salvataggio delle statistiche stimate ai dati attivi impostati per la meta - analisi con esiti continui su dati greli forniti nel dataset attivo per la stima della dimensione dell'effetto.

Studi individuali
Fornisce opzioni per il salvataggio della dimensione dell'effetto stimato.

Dimensione effetto individuale
Salva la dimensione dell'effetto stimato.

Errore standard
Salva l'errore standard stimato della dimensione dell'effetto.

Limite inferiore intervallo di confidenza
Salva l'intervallo di confidenza stimato inferiore della dimensione dell'effetto.

Limite superiore intervallo di confidenza
Salva l'intervallo di confidenza stimato per il limite superiore della dimensione dell'effetto.

Valore P
Salva il valore stimato *p*della dimensione dell'effetto.

Peso dello studio
Salva il peso dello studio stimato.

Percentuale del peso dello studio
Salva il peso dello studio normalizzato come percentuale.

Definizione Meta - Analisi Continue di salvataggio

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dati raw ...**
2. Nella finestra di dialogo **Meta - Analisi Continua** , fare clic su **Salva**.
3. Selezionare e definire le impostazioni di salvataggio delle statistiche stimate appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continua: Plot

La finestra di dialogo **Plot** fornisce le impostazioni per i seguenti tipi di plot:

- “Grafico Forest” a pagina 77
- “Grafico Forest cumulativo” a pagina 78
- “Grafico a bolle” a pagina 79
- “Grafico a imbuto” a pagina 81
- “Grafico Galbraith” a pagina 82

Meta - Analisi Continua Effetto Dimensione

La Meta - Analysis Continuous Effect Size procedura esegue la meta - analisi con esiti continui quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Esempio
Diversi studi di ricerca sono stati condotti nella storia per indagare su una medicina faddish ma discutibile per aiutare a trattare il diabete di tipo II. La medicina orale è stata pretesa di poter ridurre il livello di glucosio nel sangue dopo i pasti. I dati sono stati raccolti da diversi siti di ricerca dal 1979 al 1986.

Un investigatore principale vorrebbe trarre inferenza statistica sull'effetto della medicina orale. A causa del fatto che i dati sono stati generati da diversi studi, ha proposto l'idea di sintetizzare i risultati

attraverso gli studi per raggiungere una comprensione complessiva dell'effetto e individuare quelle sottostanti fonti di variazione degli esiti.

Statistiche
Livello di confidenza, metodo iterativo, step-halving, tolleranza di convergenza, media campionaria, varianza campionaria, deviazione standard, dimensione dell'effetto stimata, *d* di Cohen, *g* di Hedges, delta di Glass, differenza media, analisi cumulativa, metodo di stima, trim-and-fill, test basato sulla regressione, modello a effetti casuali, modello a effetti fissi, stimatore a massima verosimiglianza ristretto, stimatore di Bayes empirico, stimatore di Hedges, stimatore di Hunter-Schmidt, DerSimonian-Laird stimatore, stimatore di Sidik-Jonkman, aggiustamento dell'errore standard di Knapp-Hartung, aggiustamento dell'errore standard di Knapp-Hartung troncato, coefficienti, test di EGGER basato sulla regressione, intercetta, modello moltiplicativo, parametro di dispersione moltiplicativo, stimatore quadratico, test di omogeneità, misure di eterogeneità, intervallo di previsione, errore standard stimato, *p-value* stimato, dimensione dell'effetto complessivo cumulativo, peso stimato dello studio,

Ottenimento di un'analisi Meta - Analysis Continuous Effect Size

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dimensioni di effetto pre - calcolate**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una variabile **Effect Size** che denota la dimensione dell'effetto. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
3. Selezionare **deviazione standard** per specificare la deviazione standard dell'effetto o **Varianza** per specificare la varianza di dimensione dell'effetto e quindi selezionare una variabile per rappresentare la deviazione standard/varper la dimensione dell'effetto.
4. Facoltativamente, selezionare le variabili **Studio ID** e / o **Label di studio** . La variabile **Studio ID**selezionata non può essere la stessa della variabile selezionata **Studio Label** .
5. Facoltativamente, selezionare un'impostazione **Modello** . Quando le impostazioni **Trim - e - Fill** sono abilitate, l'impostazione controlla anche il modello utilizzato dal pooling nell'analisi del trim - e - fill. Quando le impostazioni **Bias** sono abilitate, l'impostazione controlla anche il modello utilizzato dal test basato sulla regressione.

   Effetti casuali
   L'impostazione predefinita costruisce il modello di effetti casuali.

   Effetti fissi
   Costruisce il modello a effetti fissi.
6. Facoltativamente, è possibile:
   - Clicca su **Criteri ...** per specificare i criteri generali.
   - Fare clic su **Analisi** per specificare il sottogruppo e l'analisi cumulativa.
   - Clicca su **Inferenza** per specificare i metodi di stima.
   - Clicca su **Contrast** per controllare il test di contrasto.
   - Clicca su **Bias** per accedere alla bias di pubblicazione conducendo il test di regressione di EGGER.
   - Clicca su **Trim - e - Fill** per implementare l'analisi di trim e - fill di pubblicazione bias.
   - Clicca su **Stampa** per controllare le uscite della tabella.
   - Clicca su **Salva** per salvare le statistiche stimate al dataset attivo.
   - Fare clic su **Plot** per specificare quali grafici includere nell'output.
7. Fare clic su **OK**.

Meta - Analisi Effetto Continuo Effetto: Criteri

La finestra di dialogo **Criteri** fornisce le impostazioni per specificare i criteri di meta-analisi con esiti continui quando nel dataset attivo vengono forniti i dati relativi alla dimensione dell'effetto pre - calcolato.

Intervallo di confidenza
L'impostazione opzionale specifica il livello di confidenza. Il valore deve essere un valore numerico compreso tra 0 e 100. L'impostazione predefinita è 95.

Ambito dati mancanti
Le impostazioni opzionali controllano come la procedura gestisce i dati mancanti.

Escludi casi analisi per analisi
L'impostazione predefinita include tutti i casi con dati sufficienti sulle variabili utilizzate in ogni particolare analisi.

Escludi casi listwise
Include tutti i casi con dati sufficienti su tutte le variabili utilizzate in tutte le analisi specificate dalla procedura.

Valori mancanti definiti dall'utente
Le impostazioni opzionali controllano come vengono trattati i valori mancanti dell'utente.

Escludi
L'impostazione predefinita tratta i valori mancanti dell'utente come validi.

Includi
Ignora le designazioni di valore mancanti dell'utente e tratta i valori mancanti dell'utente come validi.

Iterazioni

Numero massimo di iterazioni
L'impostazione opzionale specifica il numero massimo di iterazioni nei metodi iterativi. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore deve essere un intero intero positivo. Il valore predefinito è 100. Un valore di 0 significa che non vengono eseguite iterazioni.

Numero massimo di dimezzamenti
L'impostazione opzionale specifica il passo massimo - dimezzamento nei metodi iterativi. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore deve essere un intero intero positivo. Il valore predefinito è 5. Un valore di 0 significa che il dimezzamento non viene applicato.

Convergenza
L'impostazione opzionale specifica la tolleranza di convergenza. Il valore deve essere un valore positivo unico. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore predefinito è 1E-6.

Definizione Meta - Analisi Continua Effetto Dimensione criteri

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dimensioni di effetto pre - calcolate**
2. Nella finestra di dialogo **Meta - Analisi Continuous Effect Size** , fare clic su **Criteri**.
3. Selezionare e definire le impostazioni dei criteri appropriati.
4. Fare clic su **Continua**.

Meta - Analisi Continua Effetto Dimensione: Analisi

La finestra di dialogo **Analisi** fornisce le impostazioni per la specifica del sottogruppo e l'analisi cumulativa per la meta - analisi con esiti continui quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Analisi del sottogruppo
Selezionare una variabile che evoca l'analisi del sottogruppo. La variabile non può essere uguale alla variabile definita per **Analisi cumulativa**.

Analisi cumulativa
Seleziona una variabile che evoca l'analisi cumulativa e su cui viene condotta la meta - analisi cumulativa. La variabile non può essere uguale alla variabile definita per **Analisi Sottogruppo**. Quando **Ascending** viene selezionato, l'analisi cumulativa si basa sulla variabile specificata in ordine crescente. Quando **Descending** viene selezionato, l'analisi cumulativa si basa sulla variabile specificata in ordine decrescente.

Statistiche cumulative
Fornisce opzioni per il salvataggio delle dimensioni complessive complessive stimate. Le impostazioni sono disponibili solo quando viene selezionata una variabile **Analisi cumulativa** .

Dimensione effetto cumulativa
Salva le dimensioni complessive complessive stimate.

Errore standard
Salva l'errore standard stimato della dimensione complessiva dell'effetto cumulativo.

Limite inferiore intervallo di confidenza
Salva l'intervallo di confidenza stimato inferiore della dimensione complessiva dell'effetto cumulativo.

Limite superiore intervallo di confidenza
Salva l'intervallo di confidenza stimato per il limite superiore della dimensione complessiva dell'effetto cumulativo.

Valore P
Salva il valore stimato *p*della dimensione complessiva dell'effetto cumulativo.

Destinazione
Fornisce le opzioni per specificare un dataset di salvataggio o un file di dati. Quando **Dataset** viene selezionato, è possibile specificare un nuovo nome dataset (è possibile conservare il nome del dataset predefinito). Quando si seleziona **File di dati** , fare clic su **Sfoglia ...** per selezionare un nome file di salvataggio e la posizione.

Definizione Impostazioni di analisi di Analisi Meta - Analisi Continue Effetto

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dimensioni di effetto pre - calcolate**
2. Nella finestra di dialogo **Meta - Analysis Continuous Effect Size** , fare clic su **Analisi**.
3. Selezionare e definire le impostazioni di analisi appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continua Effetto Dimensione: Inferenza

La finestra di dialogo **Inferenza** fornisce le impostazioni per specificare i metodi di stima per la meta - analisi con esiti continui quando nel dataset attivo vengono forniti i dati relativi alla dimensione dell'effetto pre - calcolato.

Nota: La finestra di dialogo **Inferenza** è disponibile solo quando viene selezionato un modello **Random - effects** .

Stimatore
Fornisce le impostazioni per la specifica dello stimatore.

Note:

- Quando vengono specificate le impostazioni di Trim e Fill, controlla anche l'estimatore che viene utilizzato dal pooling nell'analisi del trim e del riempimento.

- Quando le impostazioni di Bias vengono specificate, controlla anche l'estimatore che viene utilizzato dal test basato sulla regressione.

Verosimiglianza massima limitata (REML)
L'impostazione predefinita applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza limitato.

Verosimiglianza massima (ML)
Applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza.

Bayes empirico
Applica il metodo iterativo e calcola lo stimatore di Bayes empirico.

Hedges
Applica il metodo non iterativo e calcola lo stimatore Hedges.

Hunter-Schmidt
Applica il metodo non iterativo e calcola lo stimatore Hunter - Schmidt.

DerSimonian-Laird
Applica il metodo non iterativo e calcola lo stimatore DerSimonian-Laird.

Sidik-Jonkman
Applica il metodo non iterativo e calcola lo stimatore Sidik-Jonkman.

Adattamento dell'errore standard
Fornisce le impostazioni che controllano se applicare la regolazione degli errori standard Knapp - Hartung.

Note:

- Quando vengono specificate le impostazioni di Trim e Fill, controlla anche la regolazione degli errori standard utilizzata dal pooling nell'analisi del trim e del riempimento.
- Quando le impostazioni di Bias vengono specificate, controlla anche la regolazione degli errori standard utilizzata dal test basato sulla regressione.

Nessun adattamento
L'impostazione predefinita non applica la regolazione.

Applica adattamento Knapp-Hartung
Applica il metodo di regolazione Knapp - Hartung.

Applica adattamento Knapp-Hartung troncato
Applica il metodo di regolazione Knapp - Hartung e tronca il valore se inferiore a 1 quando si stima la matrice di varianza - covarianza.

Definizione Meta - Analisi Continua Effetto Dimensione inferenza

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dimensioni di effetto pre - calcolate**
2. Nella finestra di dialogo **Meta - Analysis Continuous Effect Size** , clicca su **Inferenza**.
3. Selezionare e definire le impostazioni di inferenza appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Effetto Continuo Effetto: Contrasto

La finestra di dialogo **Contrast** fornisce le impostazioni per il controllo del test di contrasto per la meta - analisi con esiti continui quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Variabili
L'elenco visualizza tutte le variabili di dataset disponibili. Selezionare le variabili dall'elenco e spostarle nella lista **Contrast (s)** .

Contrasto (s)
L'elenco identifica i coefficienti che vengono memorizzati come variabili nel dataset attivo. Sono consentite più variabili. Le variabili di stringa non sono supportate.

Valori del coefficiente immesso dall'utente
Fornisce le impostazioni per la specifica dei coefficienti di un contrasto fornito dall'utente. Sono ammessi solo valori numerici. Per formulare un test di contrasto valido, il numero di valori specificati deve corrispondere a quello degli studi validi.

Definizione Meta - Analisi Continue Effetto Dimensione di contrasto

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dimensioni di effetto pre - calcolate**
2. Nella finestra di dialogo **Meta - Analysis Continuous Effect Size** , fare clic su **Contrast**.
3. Selezionare e definire le impostazioni di contrasto appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continua Effetto Dimensione: Bias

La finestra di dialogo **Bias** fornisce le impostazioni per l'abilitazione della bias di pubblicazione conducendo il test di regressione di Eggers per la meta - analisi con esiti continui quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Test basato sulla regressione di Egger
La selezione di questa impostazione abilita la bias di pubblicazione conducendo il test di regressione di Eggers.

Variabili
L'elenco fornisce tutte le variabili di dataset disponibili.

Covariate
Le variabili selezionate dall'elenco **Variabili** sono trattate come covariate. Sono consentite più covariate.

Fattori
Le variabili selezionate dall'elenco **Variabili** sono trattate come fattori. Sono consentiti più fattori.

Includi intercettazione nella regressione
Controlla il termine di intercettazione nel test di regressione.

Includi parametro di dispersione nel modello a effetti fissi
Controlla l'impostazione del modello moltiplicativo e introduce il parametro di dispersione moltiplicativo all'analisi. L'impostazione è disponibile solo quando viene selezionato un modello a effetti fissi.

Stima statistiche in base a distribuzione t
Controlla la distribuzione utilizzata nei test di regressione. L'impostazione è abilitata per impostazione predefinita, che stima le statistiche in base alla distribuzione *t*. Quando l'impostazione non viene selezionata, le statistiche vengono stimate in base alla distribuzione normale.

Definizione Meta - Analisi Continua Effetto Dimensione bias

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dimensioni di effetto pre - calcolate**
2. Nella finestra di dialogo **Meta - Analysis Continuous Effect Size** , fare clic su **Bias**.
3. Selezionare e definire le impostazioni di bias appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continua Effetto Dimensione: Trim - and - Fill

La finestra di dialogo **Trim - and - Fill** fornisce le impostazioni per implementare l'analisi di trim e - fill di pubblicazione bias per la meta - analisi con esiti continui quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Stima numero di studi mancanti
Controlla l'analisi del trim e - fill di pubblicazione bias. La selezione di questa impostazione abilita le altre impostazioni di dialogo.

Lato per assegnare gli studi
Fornisce opzioni per specificare il lato della trama del funnel su cui sono imputati gli studi mancanti.

Determinato dalla pendenza del test di Egger
L'impostazione predefinita determina il lato in base alla pendenza stimata del test di Egger.

A sinistra
Imputa il lato sinistro della trama del funnel.

Destra
Imputa il lato destro della trama del funnel.

Metodo
Specifica il metodo per stimare il numero di studi mancanti.

Lineare
L'impostazione predefinita calcola lo stimatore lineare.

Esegui
Calcola lo stimatore di esecuzione.

Quadratico
Calcola lo stimatore quadratico.

Processo di iterazione
Fornisce le impostazioni per specificare l'indicatore di iterazione e la regolazione degli errori standard.

Modello a effetti fissi
Quando viene selezionato, viene impiegato un modello a effetti fissi e la stima dell'iterazione e le opzioni di regolazione degli errori standard non sono disponibili.

Modello a effetti casuali
Quando viene selezionato, è impiegato un modello di effetti casuali e sono disponibili le seguenti impostazioni.

Stimatore
Fornisce le impostazioni per la specifica dello stimatore di iterazione

Verosimiglianza massima limitata (REML)
L'impostazione predefinita applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza limitato.

Verosimiglianza massima (ML)
Applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza.

Bayes empirico
Applica il metodo iterativo e calcola lo stimatore di Bayes empirico.

Hedges
Applica il metodo non iterativo e calcola lo stimatore Hedges.

Hunter-Schmidt
Applica il metodo non iterativo e calcola lo stimatore Hunter - Schmidt.

DerSimonian-Laird
Applica il metodo non iterativo e calcola lo stimatore DerSimonian-Laird.

Sidik-Jonkman
Applica il metodo non iterativo e calcola lo stimatore Sidik-Jonkman.

Adattamento dell'errore standard
Fornisce le impostazioni per il controllo se applicare l'adattamento di errore standard Knapp - Hartung alle iterazioni dell'algoritmo di trim e riempimento.

Nessun adattamento
L'impostazione predefinita non applica la regolazione.

Applica adattamento Knapp-Hartung
Applica il metodo di regolazione Knapp - Hartung.

Applica adattamento Knapp-Hartung troncato
Applica il metodo di regolazione Knapp - Hartung e tronca il valore se inferiore a 1 quando si stima la matrice di varianza - covarianza.

Definizione Meta - Analisi Continua Effetto Dimensione trim - e - fill

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dimensioni di effetto pre - calcolate**
2. Nella finestra di dialogo **Meta - Analysis Continuous Effect Size** , fare clic su **Trim - and - Fill**.
3. Selezionare e definire le impostazioni di allestimento e di riempimento appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continua Effetto Dimensione: Stampa

La finestra di dialogo **Stampa** fornisce le impostazioni per il controllo degli output della tabella per la meta - analisi con esiti continui quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Omogeneità/eterogeneità
Fornisce le impostazioni per il controllo dei test di omogeneità e eterogeneità.

Test di omogeneità
Quando viene selezionato, nella produzione viene fornito il corrispondente test di omogeneità.

Misure di eterogeneità
Quando selezionate, le misure di eterogeneità sono fornite nell'output.

Dimensioni effetto
Fornisce le seguenti impostazioni di dimensione dell'effetto.

Studi individuali
Controlla la visualizzazione dei singoli studi. Quando viene selezionato, viene fornita la corrispondente emissione.

Dimensioni effetto cumulativo
Controlla la visualizzazione dell'analisi cumulativa. Quando viene selezionato, viene fornita la corrispondente emissione. L'impostazione è disponibile solo quando una variabile **Analisi cumulativa** viene selezionata sulla finestra di dialogo **Analisi** .

Intervallo di previsione nel modello a effetti casuali
Controlla la visualizzazione dell'intervallo di previsione. Quando viene selezionato, viene fornita la corrispondente emissione. L'impostazione è disponibile solo quando viene specificato un modello a effetti casuali.

Definizione Meta - Analisi Continua Effetto stampa impostazioni di stampa

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dimensioni di effetto pre - calcolate**
2. Nella finestra di dialogo **Meta - Analisi Continuous Effect Size** , fare clic su **Stampa**.
3. Selezionare e definire le impostazioni di stampa appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continua Effetto Dimensione: Salva

La finestra di dialogo **Salva** fornisce le impostazioni per il salvataggio delle statistiche stimate ai dati attivi impostati per la meta - analisi con esiti continui quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Studi individuali
Fornisce opzioni per il salvataggio della dimensione dell'effetto stimato.

Errore standard
Salva l'errore standard stimato della dimensione dell'effetto. L'impostazione è disponibile solo quando non è specificata una variabile **Effect Size** .

Limite inferiore intervallo di confidenza
Salva l'intervallo di confidenza stimato inferiore della dimensione dell'effetto.

Limite superiore intervallo di confidenza
Salva l'intervallo di confidenza stimato per il limite superiore della dimensione dell'effetto.

Valore P
Salva il valore stimato *p*della dimensione dell'effetto.

Peso dello studio
Salva il peso dello studio stimato.

Percentuale del peso dello studio
Salva il peso dello studio normalizzato come percentuale.

Definizione Meta - Analisi Continua Effetto Dimensione di salvataggio

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui** > **Dimensioni di effetto pre - calcolate**
2. Nella finestra di dialogo **Meta - Analysis Continuous Effect Size** , fare clic su **Salva**.
3. Selezionare e definire le impostazioni di salvataggio delle statistiche stimate appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Continua Effetto Dimensione: Trama

La finestra di dialogo **Plot** fornisce le impostazioni per i seguenti tipi di plot:

- “Grafico Forest” a pagina 77
- “Grafico Forest cumulativo” a pagina 78
- “Grafico a bolle” a pagina 79
- “Grafico a imbuto” a pagina 81
- “Grafico Galbraith” a pagina 82

Binario meta - analisi

La Meta - Analysis Binario effettua la meta - analisi con esiti binari su dati grezzi che vengono forniti nel dataset attivo per la stima della dimensione dell'effetto.

Esempio
Diversi studi di ricerca sono stati condotti nella storia per indagare su una medicina faddish ma discutibile per aiutare a trattare il diabete di tipo II. La medicina orale è stata pretesa di poter ridurre il livello di glucosio nel sangue dopo i pasti. I dati sono stati raccolti da diversi siti di ricerca dal 1979 al 1986.

Un investigatore principale vorrebbe trarre inferenza statistica sull'effetto della medicina orale. A causa del fatto che i dati sono stati generati da diversi studi, ha proposto l'idea di sintetizzare i risultati attraverso gli studi per raggiungere una comprensione complessiva dell'effetto e individuare quelle sottostanti fonti di variazione degli esiti.

Statistiche

Intervallo di confidenza, Log Odds Ratio, Log Odds Ratio di Peto, Log Risk Ratio, differenza di rischio, effetti casuali, effetti fissi, varianza inversa, Mantel-Haenszel, iterazioni, step-halving, convergenza, statistiche cumulative, dimensione dell'effetto cumulativo, massima verosimiglianza ristretta, REML, massima verosimiglianza, ML, Empirical Bayes, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, Test di Egger, Test di Harbord, Test di Peters, intercetta nella regressione, parametro di dispersione, omogeneità, eterogeneità, statistiche esponenziali, errore standard, *valore p*, peso dello studio.

Ottenimento di un'analisi Binario Meta - Analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti Binari** > **Data raw ...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Nella sezione **Gruppo di trattamento** , selezionare una variabile **Successo** per rappresentare il "successo" conta per il gruppo di trattamento. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
3. Selezionare una variabile **Failure** per rappresentare il conteggio "guasto" per il gruppo di trattamento. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
4. Nella sezione **Gruppo di controllo** selezionare una variabile **Successo** per rappresentare il "successo" conta per il gruppo di controllo. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
5. Selezionare una variabile **Failure** per rappresentare il "fallimento" conta per il gruppo di controllo. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
6. Facoltativamente, selezionare le variabili **Studio ID** e / o **Label di studio** . La variabile **Studio ID**selezionata non può essere la stessa della variabile selezionata **Studio Label** .
7. Facoltativamente, selezionare un'impostazione **Effect Size** . Le opzioni disponibili sono **Log Odds Ratio**, **Ratio's Log Odds Ratio**, **Log Risk Ratio**e **Differenza di rischio**.
8. Facoltativamente, selezionare un'impostazione **Modello** . Quando le impostazioni **Trim - e - Fill** sono abilitate, l'impostazione controlla anche il modello utilizzato dal pooling nell'analisi del trim - e - fill. Quando le impostazioni **Bias** sono abilitate, l'impostazione controlla anche il modello utilizzato dal test basato sulla regressione.

   Effetti casuali
   L'impostazione predefinita costruisce il modello di effetti casuali.

   Effetti fissi
   Costruisce il modello a effetti fissi. **Inverse - variance** stima il peso a varianza inversa. **Mantel-Haenszel** stima il peso Mantel-Haenszel.
9. Facoltativamente, è possibile:
   - Clicca su **Criteri ...** per specificare i criteri generali.
   - Fare clic su **Analisi** per specificare il sottogruppo e l'analisi cumulativa.
   - Clicca su **Inferenza** per specificare i metodi di stima.
   - Clicca su **Contrast** per controllare il test di contrasto.
   - Clicca su **Bias** per accedere alla bias di pubblicazione conducendo il test di regressione di EGGER.
   - Clicca su **Trim - e - Fill** per implementare l'analisi di trim e - fill di pubblicazione bias.
   - Clicca su **Stampa** per controllare le uscite della tabella.
   - Clicca su **Salva** per salvare le statistiche stimate al dataset attivo.
   - Fare clic su **Plot** per specificare quali grafici includere nell'output.
10. Fare clic su **OK**.

Meta - Analisi Binario: Criteri

La finestra di dialogo **Criteri** fornisce le impostazioni per la specifica di criteri di meta-analisi con esiti binari su dati raw forniti nel dataset attivo per la stima della dimensione dell'effetto.

Intervallo di confidenza
L'impostazione opzionale specifica il livello di confidenza. Il valore deve essere un valore numerico compreso tra 0 e 100. L'impostazione predefinita è 95.

Ambito dati mancanti
Le impostazioni opzionali controllano come la procedura gestisce i dati mancanti.

Escludi casi analisi per analisi
L'impostazione predefinita include tutti i casi con dati sufficienti sulle variabili utilizzate in ogni particolare analisi.

Escludi casi listwise
Include tutti i casi con dati sufficienti su tutte le variabili utilizzate in tutte le analisi specificate dalla procedura.

Valori mancanti definiti dall'utente
Le impostazioni opzionali controllano come vengono trattati i valori mancanti dell'utente.

Escludi
L'impostazione predefinita tratta i valori mancanti dell'utente come validi.

Includi
Ignora le designazioni di valore mancanti dell'utente e tratta i valori mancanti dell'utente come validi.

Iterazioni

Numero massimo di iterazioni
L'impostazione opzionale specifica il numero massimo di iterazioni nei metodi iterativi. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore deve essere un intero intero positivo. Il valore predefinito è 100. Un valore di 0 significa che non vengono eseguite iterazioni.

Numero massimo di dimezzamenti
L'impostazione opzionale specifica il passo massimo - dimezzamento nei metodi iterativi. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore deve essere un intero intero positivo. Il valore predefinito è 5. Un valore di 0 significa che il dimezzamento non viene applicato.

Convergenza
L'impostazione opzionale specifica la tolleranza di convergenza. Il valore deve essere un valore positivo unico. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore predefinito è 1E-6.

Regola Zero Conta
Le impostazioni opzionali controllano come regolare i dati a conteggio zero, se applicabili, nella stima delle dimensioni dell'effetto.

Adatta ciascun valore solo degli studi che contengono almeno uno zero
L'impostazione predefinita regola solo i dati a conteggio zero.

Adatta ciascun valore di tutti gli studi solo se esiste almeno uno studio che contiene un valore zero
Regola tutte le frequenze quando in uno studio esiste almeno un conteggio zero.

Adatta ciascun valore di tutti gli studi
Regola tutti i dati indipendentemente dall'esistenza di zero - conta.

Non apportare correzioni e mantieni gli zeri
Non effettua alcuna regolazione dati.

Valore aggiunto
L'impostazione opzionale specifica il valore aggiunto ai dati a conteggio zero. Il valore predefinito è 0.5. Il valore specificato deve essere un valore numerico unico superiore a 0 e inferiore o uguale a 1.

Definizione Criteri Binari Meta - Analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti Binari** > **Data raw ...**
2. Nella finestra di dialogo **Meta - Analisi Binario** , fare clic su **Criteri**.
3. Selezionare e definire le impostazioni dei criteri appropriati.
4. Fare clic su **Continua**.

Meta - Analisi Binario: Analisi

La finestra di dialogo **Analisi** fornisce le impostazioni per la specifica del sottogruppo e l'analisi cumulativa per la meta - analisi con esiti binari su dati raw forniti nel dataset attivo per la stima della dimensione dell'effetto.

Analisi del sottogruppo
Selezionare una variabile che evoca l'analisi del sottogruppo. La variabile non può essere uguale alla variabile definita per **Analisi cumulativa**.

Analisi cumulativa
Seleziona una variabile che evoca l'analisi cumulativa e su cui viene condotta la meta - analisi cumulativa. La variabile non può essere uguale alla variabile definita per **Analisi Sottogruppo**. Quando **Ascending** viene selezionato, l'analisi cumulativa si basa sulla variabile specificata in ordine crescente. Quando **Descending** viene selezionato, l'analisi cumulativa si basa sulla variabile specificata in ordine decrescente.

Statistiche cumulative
Fornisce opzioni per il salvataggio delle dimensioni complessive complessive stimate. Le impostazioni sono disponibili solo quando viene selezionata una variabile **Analisi cumulativa** .

Dimensione effetto cumulativa
Salva le dimensioni complessive complessive stimate.

Dimensione effetto cumulativo (modulo espanso)
Salva le dimensioni complessive complessive stimate nella forma esponenziale.

Errore standard
Salva l'errore standard stimato della dimensione complessiva dell'effetto cumulativo.

Limite inferiore intervallo di confidenza
Salva l'intervallo di confidenza stimato inferiore della dimensione complessiva dell'effetto cumulativo.

Limite superiore intervallo di confidenza
Salva l'intervallo di confidenza stimato per il limite superiore della dimensione complessiva dell'effetto cumulativo.

Intervallo di confidenza inferiore legato (modulo espanso)
Salva l'intervallo di confidenza stimato inferiore della dimensione complessiva dell'effetto cumulativo nella forma esponenziale.

Intervallo superiore di confidenza (modulo espanso)
Salva l'intervallo di confidenza stimato, limite superiore della dimensione complessiva dell'effetto cumulativo nella forma esponenziale.

Valore P
Salva il valore stimato *p*della dimensione complessiva dell'effetto cumulativo.

Destinazione
Fornisce le opzioni per specificare un dataset di salvataggio o un file di dati. Quando **Dataset** viene selezionato, è possibile specificare un nuovo nome dataset (è possibile conservare il nome del dataset predefinito). Quando si seleziona **File di dati** , fare clic su **Sfoglia ...** per selezionare un nome file di salvataggio e la posizione.

Definizione Impostazioni di analisi di meta - analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti Binari** > **Data raw ...**
2. Nella finestra di dialogo **Meta - Analisi Binario** , fare clic su **Analisi**.
3. Selezionare e definire le impostazioni di analisi appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Binario: Inferenza

La finestra di dialogo **Inferenza** fornisce le impostazioni per specificare i metodi di stima per la meta - analisi con esiti binari su dati raw forniti nel dataset attivo per la stima della dimensione dell'effetto.

Nota: La finestra di dialogo **Inferenza** è disponibile solo quando viene selezionato un modello **Random - effects** .

Stimatore
Fornisce le impostazioni per la specifica dello stimatore.

Note:

- Quando vengono specificate le impostazioni di Trim e Fill, controlla anche l'estimatore che viene utilizzato dal pooling nell'analisi del trim e del riempimento.
- Quando le impostazioni di Bias vengono specificate, controlla anche l'estimatore che viene utilizzato dal test basato sulla regressione.

Verosimiglianza massima limitata (REML)
L'impostazione predefinita applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza limitato.

Verosimiglianza massima (ML)
Applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza.

Bayes empirico
Applica il metodo iterativo e calcola lo stimatore di Bayes empirico.

Hedges
Applica il metodo non iterativo e calcola lo stimatore Hedges.

Hunter-Schmidt
Applica il metodo non iterativo e calcola lo stimatore Hunter - Schmidt.

DerSimonian-Laird
Applica il metodo non iterativo e calcola lo stimatore DerSimonian-Laird.

Sidik-Jonkman
Applica il metodo non iterativo e calcola lo stimatore Sidik-Jonkman.

Adattamento dell'errore standard
Fornisce le impostazioni che controllano se applicare la regolazione degli errori standard Knapp - Hartung.

Note:

- Quando vengono specificate le impostazioni di Trim e Fill, controlla anche la regolazione degli errori standard utilizzata dal pooling nell'analisi del trim e del riempimento.
- Quando le impostazioni di Bias vengono specificate, controlla anche la regolazione degli errori standard utilizzata dal test basato sulla regressione.

Nessun adattamento
L'impostazione predefinita non applica la regolazione.

Applica adattamento Knapp-Hartung
Applica il metodo di regolazione Knapp - Hartung.

Applica adattamento Knapp-Hartung troncato
Applica il metodo di regolazione Knapp - Hartung e tronca il valore se inferiore a 1 quando si stima la matrice di varianza - covarianza.

Definizione Impostazioni di inferenza meta - analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti Binari** > **Data raw ...**
2. Nella finestra di dialogo **Meta - Analisi Binario** clicca su **Inferenza**.
3. Selezionare e definire le impostazioni di inferenza appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Binario: Contrasto

La finestra di dialogo **Contrast** fornisce le impostazioni per il controllo del test di contrasto per la meta - analisi con esiti binari su dati raw forniti nel dataset attivo per la stima della dimensione dell'effetto.

Variabili
L'elenco visualizza tutte le variabili di dataset disponibili. Selezionare le variabili dall'elenco e spostarle nella lista **Contrast (s)** .

Contrasto (s)
L'elenco identifica i coefficienti che vengono memorizzati come variabili nel dataset attivo. Sono consentite più variabili. Le variabili di stringa non sono supportate.

Valori del coefficiente immesso dall'utente
Fornisce le impostazioni per la specifica dei coefficienti di un contrasto fornito dall'utente. Sono ammessi solo valori numerici. Per formulare un test di contrasto valido, il numero di valori specificati deve corrispondere a quello degli studi validi.

Visualizza statistiche con esponenti
Controlla l'inclusione delle statistiche in formato esponenziale. Quando l'impostazione viene selezionata, l'output include le statistiche in formato esponenziale, comprese le dimensioni dell'effetto esponenziale e i limiti dell'intervallo di confidenza. L'impostazione è disponibile quando si specifica **Effect Size** come **Logs Odd Ratio**, **Report's Log Odd Ratio**o **Log Risk Ratio**.

Definizione Impostazioni di contrasto di meta - analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti Binari** > **Data raw ...**
2. Nella finestra di dialogo **Meta - Analisi Binario** , fare clic su **Contrast**.
3. Selezionare e definire le impostazioni di contrasto appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Binario: Bias

La finestra di dialogo **Bias** fornisce le impostazioni per l'abilitazione della bias di pubblicazione effettuando test di regressione per la meta - analisi con esiti binari su dati raw forniti nel dataset attivo per la stima della dimensione dell'effetto.

Prove a base di regressione
Fornisce le opzioni per specificare i test basati sulla regressione. Più di un test può essere selezionato.

Test di Egger
Quando selezionato, conduce il test di Egger.

Test di Harbord
Quando selezionato, conduce il test di Harbord. Il test è disponibile quando si specifica **Effect Size** come **Logs Odd Ratio** o **Log Risk Ratio**.

Test di Peters
Quando selezionato, conduce il test di Peters. Il test è disponibile quando si specifica **Effect Size** come **Logs Odd Ratio**.

Variabili
L'elenco fornisce tutte le variabili di dataset disponibili.

Covariate
Le variabili selezionate dall'elenco **Variabili** sono trattate come covariate. Sono consentite più covariate.

Fattori
Le variabili selezionate dall'elenco **Variabili** sono trattate come fattori. Sono consentiti più fattori.

Includi intercettazione nella regressione
Controlla il termine di intercettazione nel test di regressione.

Includi parametro di dispersione nel modello a effetti fissi
Controlla l'impostazione del modello moltiplicativo e introduce il parametro di dispersione moltiplicativo all'analisi. L'impostazione è disponibile solo quando viene selezionato un modello a effetti fissi.

Stima statistiche in base a distribuzione t
Controlla la distribuzione utilizzata nei test di regressione. L'impostazione è abilitata per impostazione predefinita, che stima le statistiche in base alla distribuzione *t*. Quando l'impostazione non viene selezionata, le statistiche vengono stimate in base alla distribuzione normale.

Definizione Impostazioni di bias di meta - analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti Binari** > **Data raw ...**
2. Nella finestra di dialogo **Meta - Analisi Binario** , fare clic su **Bias**.
3. Selezionare e definire le impostazioni di bias appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Binario: Trim - and - Fill

La finestra di dialogo **Trim - and - Fill** fornisce le impostazioni per implementare l'analisi trim - e - fill di pubblicazione bias per la meta - analisi con esiti binari su dati raw forniti nel dataset attivo per la stima della dimensione dell'effetto.

Stima numero di studi mancanti
Controlla l'analisi del trim e - fill di pubblicazione bias. La selezione di questa impostazione abilita le altre impostazioni di dialogo.

Lato per assegnare gli studi
Fornisce opzioni per specificare il lato della trama del funnel su cui sono imputati gli studi mancanti.

Determinato dalla pendenza del test di Egger
L'impostazione predefinita determina il lato in base alla pendenza stimata del test di Egger.

A sinistra
Imputa il lato sinistro della trama del funnel.

Destra
Imputa il lato destro della trama del funnel.

Metodo
Specifica il metodo per stimare il numero di studi mancanti.

Lineare
L'impostazione predefinita calcola lo stimatore lineare.

Esegui
Calcola lo stimatore di esecuzione.

Quadratico
Calcola lo stimatore quadratico.

Processo di iterazione
Fornisce le impostazioni per specificare l'indicatore di iterazione e la regolazione degli errori standard.

Modello a effetti fissi
Quando viene selezionato, viene impiegato un modello a effetti fissi e la stima dell'iterazione e le opzioni di regolazione degli errori standard non sono disponibili.

Modello a effetti casuali
Quando viene selezionato, è impiegato un modello di effetti casuali e sono disponibili le seguenti impostazioni.

Stimatore
Fornisce le impostazioni per la specifica dello stimatore di iterazione

Verosimiglianza massima limitata (REML)
L'impostazione predefinita applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza limitato.

Verosimiglianza massima (ML)
Applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza.

Bayes empirico
Applica il metodo iterativo e calcola lo stimatore di Bayes empirico.

Hedges
Applica il metodo non iterativo e calcola lo stimatore Hedges.

Hunter-Schmidt
Applica il metodo non iterativo e calcola lo stimatore Hunter - Schmidt.

DerSimonian-Laird
Applica il metodo non iterativo e calcola lo stimatore DerSimonian-Laird.

Sidik-Jonkman
Applica il metodo non iterativo e calcola lo stimatore Sidik-Jonkman.

Adattamento dell'errore standard
Fornisce le impostazioni per il controllo se applicare l'adattamento di errore standard Knapp - Hartung alle iterazioni dell'algoritmo di trim e riempimento.

Nessun adattamento
L'impostazione predefinita non applica la regolazione.

Applica adattamento Knapp-Hartung
Applica il metodo di regolazione Knapp - Hartung.

Applica adattamento Knapp-Hartung troncato
Applica il metodo di regolazione Knapp - Hartung e tronca il valore se inferiore a 1 quando si stima la matrice di varianza - covarianza.

Definizione Meta - Analisi Binario di trim - e - fill

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti Binari** > **Data raw ...**
2. Nella finestra di dialogo **Meta - Analisi Binario** , fare clic su **Trim - and - Fill**.
3. Selezionare e definire le impostazioni di allestimento e di riempimento appropriate.

4. Fare clic su **Continua**.

Meta - Analisi Binario: Stampa

La finestra di dialogo **Stampa** fornisce le impostazioni per il controllo degli output della tabella per la meta - analisi con esiti binari su dati greli forniti nel dataset attivo per la stima della dimensione dell'effetto.

Omogeneità/eterogeneità
Fornisce le impostazioni per il controllo dei test di omogeneità e eterogeneità.

Test di omogeneità
Quando viene selezionato, nella produzione viene fornito il corrispondente test di omogeneità.

Misure di eterogeneità
Quando selezionate, le misure di eterogeneità sono fornite nell'output.

Dimensioni effetto
Fornisce le seguenti impostazioni di dimensione dell'effetto.

Studi individuali
Controlla la visualizzazione dei singoli studi. Quando viene selezionato, viene fornita la corrispondente emissione.

Dimensioni effetto cumulativo
Controlla la visualizzazione dell'analisi cumulativa. Quando viene selezionato, viene fornita la corrispondente emissione. L'impostazione è disponibile solo quando una variabile **Analisi cumulativa** viene selezionata sulla finestra di dialogo **Analisi** .

Intervallo di previsione nel modello a effetti casuali
Controlla la visualizzazione dell'intervallo di previsione. Quando viene selezionato, viene fornita la corrispondente emissione. L'impostazione è disponibile solo quando viene specificato un modello a effetti casuali.

Visualizza statistiche con esponenti
Controlla l'inclusione delle statistiche in formato esponenziale. Quando l'impostazione viene selezionata, vengono emesse le statistiche in formato esponenziale, comprese le dimensioni dell'effetto esponenziale e i limiti dell'intervallo di confidenza. Il processo si applica sia alle tabelle di stima e di previsione degli effetti sia obbligatorie che facoltative. L'impostazione è disponibile quando si specifica **Effect Size** come **Logs Odd Ratio**, **Report's Log Odd Ratio**o **Log Risk Ratio**.

Definizione Impostazioni di stampa di meta - analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti Binari** > **Data raw ...**
2. Nella finestra di dialogo **Meta - Analisi Binario** , fare clic su **Stampa**.
3. Selezionare e definire le impostazioni di stampa appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Binario: Salva

La finestra di dialogo **Salva** fornisce le impostazioni per il salvataggio delle statistiche stimate ai dati attivi impostati per la meta - analisi con esiti binari su dati greli forniti nel dataset attivo per la stima della dimensione dell'effetto.

Studi individuali
Fornisce opzioni per il salvataggio della dimensione dell'effetto stimato.

Dimensione effetto individuale
Salva la dimensione dell'effetto stimato.

Dimensione effetto individuale (modulo esponenziale)
Salva la dimensione dell'effetto stimato nella forma esponenziale.

Errore standard
Salva l'errore standard stimato della dimensione dell'effetto.

Limite inferiore intervallo di confidenza
Salva l'intervallo di confidenza stimato inferiore della dimensione dell'effetto.

Limite superiore intervallo di confidenza
Salva l'intervallo di confidenza stimato per il limite superiore della dimensione dell'effetto.

Intervallo di confidenza inferiore legato (modulo espanso)
Salva l'intervallo di confidenza stimato inferiore della dimensione dell'effetto nella forma esponenziale.

Intervallo superiore di confidenza (modulo espanso)
Salva l'intervallo di confidenza stimato dell'intervallo di confidenza della dimensione dell'effetto nella forma esponenziale.

Valore P
Salva il valore stimato *p*della dimensione dell'effetto.

Peso dello studio
Salva il peso dello studio stimato.

Percentuale del peso dello studio
Salva il peso dello studio normalizzato come percentuale.

Definizione Meta - Analisi Binario di salvataggio

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti Binari** > **Data raw ...**
2. Nella finestra di dialogo **Meta - Analisi Binario** , fare clic su **Salva**.
3. Selezionare e definire le impostazioni di salvataggio delle statistiche stimate appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Binario: Plot

La finestra di dialogo **Plot** fornisce le impostazioni per i seguenti tipi di plot:

- “Grafico Forest” a pagina 77
- “Grafico Forest cumulativo” a pagina 78
- “Grafico a bolle” a pagina 79
- “Grafico a imbuto” a pagina 81
- “Grafico Galbraith” a pagina 82
- “L'Abb'e Plot” a pagina 82

Meta - Analisi Effetto Binario Dimensione

La Meta - Analysis Binary Effect Size procedura esegue la meta - analisi con esiti binari quando nel dataset attivo vengono forniti i dati relativi alla dimensione dell'effetto pre - calcolato.

Esempio
Diversi studi di ricerca sono stati condotti nella storia per indagare su una medicina faddish ma discutibile per aiutare a trattare il diabete di tipo II. La medicina orale è stata pretesa di poter ridurre il livello di glucosio nel sangue dopo i pasti. I dati sono stati raccolti da diversi siti di ricerca dal 1979 al 1986.

Un investigatore principale vorrebbe trarre inferenza statistica sull'effetto della medicina orale. A causa del fatto che i dati sono stati generati da diversi studi, ha proposto l'idea di sintetizzare i risultati attraverso gli studi per raggiungere una comprensione complessiva dell'effetto e individuare quelle sottostanti fonti di variazione degli esiti.

Statistiche

Intervallo di confidenza, Log Odds Ratio, Log Odds Ratio di Peto, Log Risk Ratio, differenza di rischio, effetti casuali, effetti fissi, varianza inversa, Mantel-Haenszel, iterazioni, step-halving, convergenza, statistiche cumulative, dimensione dell'effetto cumulativo, massima verosimiglianza ristretta, REML, massima verosimiglianza, ML, Empirical Bayes, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, Test di Egger, Test di Harbord, Test di Peters, intercetta nella regressione, parametro di dispersione, omogeneità, eterogeneità, statistiche esponenziali, errore standard, *valore p*, peso dello studio.

Ottenimento di un'analisi Meta - Analisi Binario Effetto Dimensione

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti binari** > **Dimensioni di effetto pre - calcolate ...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una variabile **Effect Size** che denota la dimensione dell'effetto. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
3. Selezionare **deviazione standard** per specificare la deviazione standard dell'effetto o **Varianza** per specificare la varianza di dimensione dell'effetto e quindi selezionare una variabile per rappresentare la deviazione standard/varper la dimensione dell'effetto.
4. Facoltativamente, selezionare le variabili **Studio ID** e / o **Label di studio** . La variabile **Studio ID**selezionata non può essere la stessa della variabile selezionata **Studio Label** .
5. Facoltativamente, selezionare un'impostazione **Effect Size** . Le opzioni disponibili sono **Log Odds Ratio**, **Ratio's Log Odds Ratio**, **Log Risk Ratio**e **Differenza di rischio**.
6. Facoltativamente, selezionare un'impostazione **Modello** . Quando le impostazioni **Trim - e - Fill** sono abilitate, l'impostazione controlla anche il modello utilizzato dal pooling nell'analisi del trim - e - fill. Quando le impostazioni **Bias** sono abilitate, l'impostazione controlla anche il modello utilizzato dal test basato sulla regressione.

   Effetti casuali

   L'impostazione predefinita costruisce il modello di effetti casuali.

   Effetti fissi

   Costruisce il modello a effetti fissi.
7. Facoltativamente, è possibile:
   - Clicca su **Criteri ...** per specificare i criteri generali.
   - Fare clic su **Analisi** per specificare il sottogruppo e l'analisi cumulativa.
   - Clicca su **Inferenza** per specificare i metodi di stima.
   - Clicca su **Contrast** per controllare il test di contrasto.
   - Clicca su **Bias** per accedere alla bias di pubblicazione conducendo il test di regressione di EGGER.
   - Clicca su **Trim - e - Fill** per implementare l'analisi di trim e - fill di pubblicazione bias.
   - Clicca su **Stampa** per controllare le uscite della tabella.
   - Clicca su **Salva** per salvare le statistiche stimate al dataset attivo.
   - Fare clic su **Plot** per specificare quali grafici includere nell'output.
8. Fare clic su **OK**.

Meta - Analisi Effetto Binario Dimensione: Criteri

La finestra di dialogo **Criteri** fornisce le impostazioni per la specifica di criteri di meta - analisi con esiti binari quando nel dataset attivo vengono forniti i dati relativi alla dimensione dell'effetto pre - calcolato.

Intervallo di confidenza
L'impostazione opzionale specifica il livello di confidenza. Il valore deve essere un valore numerico compreso tra 0 e 100. L'impostazione predefinita è 95.

Ambito dati mancanti
Le impostazioni opzionali controllano come la procedura gestisce i dati mancanti.

Escludi casi analisi per analisi
L'impostazione predefinita include tutti i casi con dati sufficienti sulle variabili utilizzate in ogni particolare analisi.

Escludi casi listwise
Include tutti i casi con dati sufficienti su tutte le variabili utilizzate in tutte le analisi specificate dalla procedura.

Valori mancanti definiti dall'utente
Le impostazioni opzionali controllano come vengono trattati i valori mancanti dell'utente.

Escludi
L'impostazione predefinita tratta i valori mancanti dell'utente come validi.

Includi
Ignora le designazioni di valore mancanti dell'utente e tratta i valori mancanti dell'utente come validi.

Iterazioni

Numero massimo di iterazioni
L'impostazione opzionale specifica il numero massimo di iterazioni nei metodi iterativi. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore deve essere un intero intero positivo. Il valore predefinito è 100. Un valore di 0 significa che non vengono eseguite iterazioni.

Numero massimo di dimezzamenti
L'impostazione opzionale specifica il passo massimo - dimezzamento nei metodi iterativi. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore deve essere un intero intero positivo. Il valore predefinito è 5. Un valore di 0 significa che il dimezzamento non viene applicato.

Convergenza
L'impostazione opzionale specifica la tolleranza di convergenza. Il valore deve essere un valore positivo unico. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore predefinito è 1E-6.

Definizione Meta - Analisi Effetto Binario criteri binari

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti binari** > **Dimensioni di effetto pre - calcolate ...**
2. Nella finestra di dialogo **Meta - Analysis Binary Effect Size** , fare clic su **Criteri**.
3. Selezionare e definire le impostazioni dei criteri appropriati.
4. Fare clic su **Continua**.

Meta - Analisi Effetto Binario Effetto: Analisi

La finestra di dialogo **Analisi** fornisce le impostazioni per la specifica del sottogruppo e l'analisi cumulativa per la meta - analisi con esiti binari quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Analisi del sottogruppo
Selezionare una variabile che evoca l'analisi del sottogruppo. La variabile non può essere uguale alla variabile definita per **Analisi cumulativa**.

Analisi cumulativa
Seleziona una variabile che evoca l'analisi cumulativa e su cui viene condotta la meta - analisi cumulativa. La variabile non può essere uguale alla variabile definita per **Analisi Sottogruppo**. Quando **Ascending** viene selezionato, l'analisi cumulativa si basa sulla variabile specificata in ordine crescente. Quando **Descending** viene selezionato, l'analisi cumulativa si basa sulla variabile specificata in ordine decrescente.

Statistiche cumulative
Fornisce opzioni per il salvataggio delle dimensioni complessive complessive stimate. Le impostazioni sono disponibili solo quando viene selezionata una variabile **Analisi cumulativa** .

Dimensione effetto cumulativa
Salva le dimensioni complessive complessive stimate.

Dimensione effetto cumulativo (modulo espanso)
Salva le dimensioni complessive complessive stimate nella forma esponenziale.

Errore standard
Salva l'errore standard stimato della dimensione complessiva dell'effetto cumulativo.

Limite inferiore intervallo di confidenza
Salva l'intervallo di confidenza stimato inferiore della dimensione complessiva dell'effetto cumulativo.

Limite superiore intervallo di confidenza
Salva l'intervallo di confidenza stimato per il limite superiore della dimensione complessiva dell'effetto cumulativo.

Intervallo di confidenza inferiore legato (modulo espanso)
Salva l'intervallo di confidenza stimato inferiore della dimensione complessiva dell'effetto cumulativo nella forma esponenziale.

Intervallo superiore di confidenza (modulo espanso)
Salva l'intervallo di confidenza stimato, limite superiore della dimensione complessiva dell'effetto cumulativo nella forma esponenziale.

Valore P
Salva il valore stimato *p*della dimensione complessiva dell'effetto cumulativo.

Destinazione
Fornisce le opzioni per specificare un dataset di salvataggio o un file di dati. Quando **Dataset** viene selezionato, è possibile specificare un nuovo nome dataset (è possibile conservare il nome del dataset predefinito). Quando si seleziona **File di dati** , fare clic su **Sfoglia ...** per selezionare un nome file di salvataggio e la posizione.

Definizione Impostazioni di analisi di Analisi Meta - Analisi Effetto

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti binari** > **Dimensioni di effetto pre - calcolate ...**
2. Nella finestra di dialogo **Meta - Analysis Binary Effect Size** , fare clic su **Analisi**.
3. Selezionare e definire le impostazioni di analisi appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Binario Effetto Binario: Inferenza

La finestra di dialogo **Inferenza** fornisce le impostazioni per specificare i metodi di stima per la meta - analisi con esiti binari quando nel dataset attivo vengono forniti i dati relativi alla dimensione dell'effetto pre - calcolato.

Stimatore
Fornisce le impostazioni per la specifica dello stimatore.

Note:

- Quando vengono specificate le impostazioni di Trim e Fill, controlla anche l'estimatore che viene utilizzato dal pooling nell'analisi del trim e del riempimento.
- Quando le impostazioni di Bias vengono specificate, controlla anche l'estimatore che viene utilizzato dal test basato sulla regressione.

Verosimiglianza massima limitata (REML)
L'impostazione predefinita applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza limitato.

Verosimiglianza massima (ML)
Applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza.

Bayes empirico
Applica il metodo iterativo e calcola lo stimatore di Bayes empirico.

Hedges
Applica il metodo non iterativo e calcola lo stimatore Hedges.

Hunter-Schmidt
Applica il metodo non iterativo e calcola lo stimatore Hunter - Schmidt.

DerSimonian-Laird
Applica il metodo non iterativo e calcola lo stimatore DerSimonian-Laird.

Sidik-Jonkman
Applica il metodo non iterativo e calcola lo stimatore Sidik-Jonkman.

Adattamento dell'errore standard
Fornisce le impostazioni che controllano se applicare la regolazione degli errori standard Knapp - Hartung.

Note:

- Quando vengono specificate le impostazioni di Trim e Fill, controlla anche la regolazione degli errori standard utilizzata dal pooling nell'analisi del trim e del riempimento.
- Quando le impostazioni di Bias vengono specificate, controlla anche la regolazione degli errori standard utilizzata dal test basato sulla regressione.

Nessun adattamento
L'impostazione predefinita non applica la regolazione.

Applica adattamento Knapp-Hartung
Applica il metodo di regolazione Knapp - Hartung.

Applica adattamento Knapp-Hartung troncato
Applica il metodo di regolazione Knapp - Hartung e tronca il valore se inferiore a 1 quando si stima la matrice di varianza - covarianza.

Definizione Meta - Analisi Effetto Binario Effetto delle impostazioni di inferenza

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti binari** > **Dimensioni di effetto pre - calcolate ...**
2. Nella finestra di dialogo **Meta - Analysis Binary Effect Size** , clicca su **Inferenza**.
3. Selezionare e definire le impostazioni di inferenza appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Effetto Binario Effetto: Contrasto

La finestra di dialogo **Contrast** fornisce le impostazioni per il controllo del test di contrasto per la meta - analisi con esiti binari quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Variabili
L'elenco visualizza tutte le variabili di dataset disponibili. Selezionare le variabili dall'elenco e spostarle nella lista **Contrast (s)** .

Contrasto (s)
L'elenco identifica i coefficienti che vengono memorizzati come variabili nel dataset attivo. Sono consentite più variabili. Le variabili di stringa non sono supportate.

Valori del coefficiente immesso dall'utente
Fornisce le impostazioni per la specifica dei coefficienti di un contrasto fornito dall'utente. Sono ammessi solo valori numerici. Per formulare un test di contrasto valido, il numero di valori specificati deve corrispondere a quello degli studi validi.

Visualizza statistiche con esponenti
Controlla l'inclusione delle statistiche in formato esponenziale. Quando l'impostazione viene selezionata, l'output include le statistiche in formato esponenziale, comprese le dimensioni dell'effetto esponenziale e i limiti dell'intervallo di confidenza. L'impostazione è disponibile quando si specifica **Effect Size** come **Logs Odd Ratio**, **Report's Log Odd Ratio**o **Log Risk Ratio**.

Definizione Meta - Analisi Effetto Binario Effetto di contrasto

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti binari** > **Dimensioni di effetto pre - calcolate ...**
2. Nella finestra di dialogo **Meta - Analysis Binary Effect Size** , fare clic su **Contrast**.
3. Selezionare e definire le impostazioni di contrasto appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Effetto Binario Effetto: Bias

La finestra di dialogo **Bias** fornisce le impostazioni per l'abilitazione della bias di pubblicazione effettuando test di regressione per la meta - analisi con esiti binari quando nel dataset attivo vengono forniti i dati relativi alla dimensione dell'effetto pre - calcolato.

Test basato sulla regressione di Egger
La selezione di questa impostazione abilita la bias di pubblicazione conducendo il test di regressione di Eggers.

Variabili
L'elenco fornisce tutte le variabili di dataset disponibili.

Covariate
Le variabili selezionate dall'elenco **Variabili** sono trattate come covariate. Sono consentite più covariate.

Fattori
Le variabili selezionate dall'elenco **Variabili** sono trattate come fattori. Sono consentiti più fattori.

Includi intercettazione nella regressione
Controlla il termine di intercettazione nel test di regressione.

Includi parametro di dispersione nel modello a effetti fissi
Controlla l'impostazione del modello moltiplicativo e introduce il parametro di dispersione moltiplicativo all'analisi. L'impostazione è disponibile solo quando viene selezionato un modello a effetti fissi.

Stima statistiche in base a distribuzione t
Controlla la distribuzione utilizzata nei test di regressione. L'impostazione è abilitata per impostazione predefinita, che stima le statistiche in base alla distribuzione *t*. Quando l'impostazione non viene selezionata, le statistiche vengono stimate in base alla distribuzione normale.

Definizione Meta - Analisi Effetto Binario Dimensione impostazioni bias

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti binari** > **Dimensioni di effetto pre - calcolate ...**
2. Nella finestra di dialogo **Meta - Analysis Binary Effect Size** , fare clic su **Bias**.

3. Selezionare e definire le impostazioni di bias appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Effetto Binario Effetto: Trim - e - Fill

La finestra di dialogo **Trim - and - Fill** fornisce le impostazioni per implementare l'analisi di trim e - fill di pubblicazione bias per la meta - analisi con esiti binari quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Stima numero di studi mancanti
Controlla l'analisi del trim e - fill di pubblicazione bias. La selezione di questa impostazione abilita le altre impostazioni di dialogo.

Lato per assegnare gli studi
Fornisce opzioni per specificare il lato della trama del funnel su cui sono imputati gli studi mancanti.

Determinato dalla pendenza del test di Egger
L'impostazione predefinita determina il lato in base alla pendenza stimata del test di Egger.

A sinistra
Imputa il lato sinistro della trama del funnel.

Destra
Imputa il lato destro della trama del funnel.

Metodo
Specifica il metodo per stimare il numero di studi mancanti.

Lineare
L'impostazione predefinita calcola lo stimatore lineare.

Esegui
Calcola lo stimatore di esecuzione.

Quadratico
Calcola lo stimatore quadratico.

Processo di iterazione
Fornisce le impostazioni per specificare l'indicatore di iterazione e la regolazione degli errori standard.

Modello a effetti fissi
Quando viene selezionato, viene impiegato un modello a effetti fissi e la stima dell'iterazione e le opzioni di regolazione degli errori standard non sono disponibili.

Modello a effetti casuali
Quando viene selezionato, è impiegato un modello di effetti casuali e sono disponibili le seguenti impostazioni.

Stimatore
Fornisce le impostazioni per la specifica dello stimatore di iterazione

Verosimiglianza massima limitata (REML)
L'impostazione predefinita applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza limitato.

Verosimiglianza massima (ML)
Applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza.

Bayes empirico
Applica il metodo iterativo e calcola lo stimatore di Bayes empirico.

Hedges
Applica il metodo non iterativo e calcola lo stimatore Hedges.

Hunter-Schmidt
Applica il metodo non iterativo e calcola lo stimatore Hunter - Schmidt.

DerSimonian-Laird
Applica il metodo non iterativo e calcola lo stimatore DerSimonian-Laird.

Sidik-Jonkman
Applica il metodo non iterativo e calcola lo stimatore Sidik-Jonkman.

Adattamento dell'errore standard
Fornisce le impostazioni per il controllo se applicare l'adattamento di errore standard Knapp - Hartung alle iterazioni dell'algoritmo di trim e riempimento.

Nessun adattamento
L'impostazione predefinita non applica la regolazione.

Applica adattamento Knapp-Hartung
Applica il metodo di regolazione Knapp - Hartung.

Applica adattamento Knapp-Hartung troncato
Applica il metodo di regolazione Knapp - Hartung e tronca il valore se inferiore a 1 quando si stima la matrice di varianza - covarianza.

Definizione Meta - Analisi Effetto Binario Dimensione trim - e - fill

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti binari** > **Dimensioni di effetto pre - calcolate ...**
2. Nella finestra di dialogo **Meta - Analysis Binary Effect Size** , clicca su **Trim - and - Fill**.
3. Selezionare e definire le impostazioni di allestimento e di riempimento appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Effetto Binario Effetto: Stampa

La finestra di dialogo **Stampa** fornisce le impostazioni per il controllo degli output della tabella per la meta - analisi con esiti binari quando nel dataset attivo vengono forniti i dati relativi alla dimensione dell'effetto pre - calcolato.

Omogeneità/eterogeneità
Fornisce le impostazioni per il controllo dei test di omogeneità e eterogeneità.

Test di omogeneità
Quando viene selezionato, nella produzione viene fornito il corrispondente test di omogeneità.

Misure di eterogeneità
Quando selezionate, le misure di eterogeneità sono fornite nell'output.

Dimensioni effetto
Fornisce le seguenti impostazioni di dimensione dell'effetto.

Studi individuali
Controlla la visualizzazione dei singoli studi. Quando viene selezionato, viene fornita la corrispondente emissione.

Dimensioni effetto cumulativo
Controlla la visualizzazione dell'analisi cumulativa. Quando viene selezionato, viene fornita la corrispondente emissione. L'impostazione è disponibile solo quando una variabile **Analisi cumulativa** viene selezionata sulla finestra di dialogo **Analisi** .

Intervallo di previsione nel modello a effetti casuali
Controlla la visualizzazione dell'intervallo di previsione. Quando viene selezionato, viene fornita la corrispondente emissione. L'impostazione è disponibile solo quando viene specificato un modello a effetti casuali.

Visualizza statistiche con esponenti
Controlla l'inclusione delle statistiche in formato esponenziale. Quando l'impostazione viene selezionata, vengono emesse le statistiche in formato esponenziale, comprese le dimensioni dell'effetto esponenziale e i limiti dell'intervallo di confidenza. Il processo si applica sia alle tabelle di stima e di previsione degli effetti sia obbligatorie che facoltative. L'impostazione è disponibile quando si specifica **Effect Size** come **Logs Odd Ratio**, **Report's Log Odd Ratio**o **Log Risk Ratio**.

Definizione Impostazioni di stampa di Analisi Meta - Analisi Effetto

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti binari** > **Dimensioni di effetto pre - calcolate ...**
2. Nella finestra di dialogo **Meta - Analysis Binary Effect Size** , fare clic su **Stampa**.
3. Selezionare e definire le impostazioni di stampa appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Effetto Binario Effetto: Salva

La finestra di dialogo **Salva** fornisce le impostazioni per il salvataggio delle statistiche stimate ai dati attivi impostati per la meta - analisi con esiti binari quando i dati di dimensione dell'effetto pre - calcolati sono forniti nel dataset attivo.

Studi individuali
Fornisce opzioni per il salvataggio della dimensione dell'effetto stimato.

Dimensione effetto individuale (modulo esponenziale)
Salva la dimensione dell'effetto stimato nella forma esponenziale.

Errore standard
Salva l'errore standard stimato della dimensione dell'effetto.

Limite inferiore intervallo di confidenza
Salva l'intervallo di confidenza stimato inferiore della dimensione dell'effetto.

Limite superiore intervallo di confidenza
Salva l'intervallo di confidenza stimato per il limite superiore della dimensione dell'effetto.

Intervallo di confidenza inferiore legato (modulo espanso)
Salva l'intervallo di confidenza stimato inferiore della dimensione dell'effetto nella forma esponenziale.

Intervallo superiore di confidenza (modulo espanso)
Salva l'intervallo di confidenza stimato dell'intervallo di confidenza della dimensione dell'effetto nella forma esponenziale.

Valore P
Salva il valore stimato *p*della dimensione dell'effetto.

Peso dello studio
Salva il peso dello studio stimato.

Percentuale del peso dello studio
Salva il peso dello studio normalizzato come percentuale.

Definizione Meta - Analisi Effetto Binario Dimensione risparmio impostazioni

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti binari** > **Dimensioni di effetto pre - calcolate ...**
2. Nella finestra di dialogo **Meta - Analysis Binary Effect Size** , fare clic su **Salva**.
3. Selezionare e definire le impostazioni di salvataggio delle statistiche stimate appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Effetto Binario Dimensione: Trama

La finestra di dialogo **Plot** fornisce le impostazioni per i seguenti tipi di plot:

- “Grafico Forest” a pagina 77
- “Grafico Forest cumulativo” a pagina 78
- “Grafico a bolle” a pagina 79

- “Grafico a imbuto” a pagina 81
- “Grafico Galbraith” a pagina 82
- “L'Abb'e Plot” a pagina 82

Regressione meta - analisi

La Procedura Meta - Analysis Regression esegue analisi di meta - regressione.

Esempio

Diversi studi di ricerca sono stati condotti nella storia per indagare su una medicina faddish ma discutibile per aiutare a trattare il diabete di tipo II. La medicina orale è stata pretesa di poter ridurre il livello di glucosio nel sangue dopo i pasti. I dati sono stati raccolti da diversi siti di ricerca dal 1979 al 1986.

Un investigatore principale vorrebbe trarre inferenza statistica sull'effetto della medicina orale. A causa del fatto che i dati sono stati generati da diversi studi, ha proposto l'idea di sintetizzare i risultati attraverso gli studi per raggiungere una comprensione complessiva dell'effetto e individuare quelle sottostanti fonti di variazione degli esiti.

Statistiche

Livello di confidenza, metodo iterativo, step-halving, tolleranza di convergenza, media campionaria, varianza campionaria, deviazione standard, dimensione dell'effetto stimata, metodo di stima, test basato sulla regressione, modello a effetti casuali, modello a effetti fissi, parametro di dispersione, stimatore di massima verosimiglianza ristretto, stimatore empirico di Bayes, stimatore di Hedges, stimatore di Hunter-Schmidt, stimatore di DerSimonian-Laird, stimatore di Sidik-Jonkman, aggiustamento dell'errore standard di Knapp-Hartung, aggiustamento dell'errore standard di Knapp-Hartung troncato, coefficienti, test dei coefficienti del modello, statistiche esponenziate.

Ottenimento di un'analisi Meta - Analisi Regression

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Meta Regressione**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una variabile singola dipendente **Effect Size** che denota la dimensione dell'effetto. La variabile selezionata deve essere numerica (le variabili stringa non sono supportate).
3. Selezionare una delle seguenti impostazioni e quindi selezionare una variabile numerica singola corrispondente:

   Errore standard
   Selezionare una variabile che specifica l'errore standard convertito in peso. Questa è l'impostazione predefinita.

   Varianza
   Selezionare una variabile che specifica la varianza che viene convertita in peso.

   Peso
   Selezionare una variabile che specifica il peso.
4. Facoltativamente, aggiungere le variabili del fattore all'elenco **Fattore (s)** . Per ogni variabile fattore selezionata, è possibile specificare un valore opzionale per ogni variabile per designare **Ultime categorie**personalizzate. È possibile fare clic su **Reset** per ripristinare i valori **Ultime categorie** alle relative impostazioni predefinite.

   Nota: Quando non ci sono casi che corrispondono ai valori specificati **Ultime categorie** , gli ultimi valori che si verificano sono trattati come le ultime categorie.
5. Facoltativamente, selezionare variabili covariate numeriche.
6. Facoltativamente, selezionare un'impostazione **Modello** .

Effetti casuali
L'impostazione predefinita costruisce il modello di effetti casuali.

Effetti fissi
Costruisce il modello a effetti fissi. È possibile opzionalmente selezionare l'impostazione **Includi parametro di dispersione** .

7. Facoltativamente, è possibile:
   - Clicca su **Criteri ...** per specificare i criteri generali.
   - Clicca su **Inferenza** per specificare i metodi di stima.
   - Clicca su **Stampa** per controllare le uscite della tabella.
   - Clicca su **Salva** per prevedere e salvare le statistiche stimate al dataset attivo.
   - Fare clic su **Plot** per specificare quali grafici includere nell'output.
8. Fare clic su **OK**.

Regressione Meta - Analisi: Criteri

La finestra di dialogo **Criteri** fornisce le impostazioni per specificare i criteri per la regressione meta-analisi.

Intervallo di confidenza
L'impostazione opzionale specifica il livello di confidenza. Il valore deve essere un valore numerico compreso tra 0 e 100. L'impostazione predefinita è 95.

Valori mancanti definiti dall'utente
Le impostazioni opzionali controllano come vengono trattati i valori mancanti dell'utente.

Escludi
L'impostazione predefinita tratta i valori mancanti dell'utente come validi.

Includi
Ignora le designazioni di valore mancanti dell'utente e tratta i valori mancanti dell'utente come validi.

Iterazioni

Numero massimo di iterazioni
L'impostazione opzionale specifica il numero massimo di iterazioni nei metodi iterativi. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore deve essere un intero intero positivo. Il valore predefinito è 100. Un valore di 0 significa che non vengono eseguite iterazioni.

Numero massimo di dimezzamenti
L'impostazione opzionale specifica il passo massimo - dimezzamento nei metodi iterativi. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore deve essere un intero intero positivo. Il valore predefinito è 5. Un valore di 0 significa che il dimezzamento non viene applicato.

Convergenza
L'impostazione opzionale specifica la tolleranza di convergenza. Il valore deve essere un valore positivo unico. L'impostazione è disponibile quando vengono utilizzati metodi iterativi. Il valore predefinito è 1E-6.

Definizione Criteri Di Regressione Meta - Analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Meta Regressione**
2. Nella finestra di dialogo **Meta - Analysis Regression** , fare clic su **Criteri**.
3. Selezionare e definire le impostazioni dei criteri appropriati.
4. Fare clic su **Continua**.

Regressione meta - analisi: Inferenza

La finestra di dialogo **Inferenza** fornisce le impostazioni per specificare i metodi di stima per la regressione meta-analisi.

Includi intercettazione nella regressione
L'impostazione è abilitata essere predefinita.

Stima statistiche in base a distribuzione t
Controlla la distribuzione utilizzata nei test di regressione. L'impostazione è abilitata per impostazione predefinita, che stima le statistiche in base alla distribuzione *t*. Quando l'impostazione non viene selezionata, le statistiche vengono stimate in base alla distribuzione normale.

Stimatore
Fornisce le impostazioni per la specifica dello stimatore.

Verosimiglianza massima limitata (REML)
L'impostazione predefinita applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza limitato.

Verosimiglianza massima (ML)
Applica il metodo iterativo e calcola lo stimatore di massima verosimiglianza.

Bayes empirico
Applica il metodo iterativo e calcola lo stimatore di Bayes empirico.

Hedges
Applica il metodo non iterativo e calcola lo stimatore Hedges.

Hunter-Schmidt
Applica il metodo non iterativo e calcola lo stimatore Hunter - Schmidt.

DerSimonian-Laird
Applica il metodo non iterativo e calcola lo stimatore DerSimonian-Laird.

Sidik-Jonkman
Applica il metodo non iterativo e calcola lo stimatore Sidik-Jonkman.

Adattamento dell'errore standard
Fornisce le impostazioni che controllano se applicare la regolazione degli errori standard Knapp - Hartung.

Nessun adattamento
L'impostazione predefinita non applica la regolazione.

Applica adattamento Knapp-Hartung
Applica il metodo di regolazione Knapp - Hartung.

Applica adattamento Knapp-Hartung troncato
Applica il metodo di regolazione Knapp - Hartung e tronca il valore se inferiore a 1 quando si stima la matrice di varianza - covarianza.

Definizione Impostazioni di inferenza di regressione meta - analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Meta Regressione**
2. Nella finestra di dialogo **Meta - Analysis Regression** , fare clic su **Inferenza**.
3. Selezionare e definire le impostazioni di inferenza appropriate.
4. Fare clic su **Continua**.

Regressione meta - analisi: Stampa

La finestra di dialogo **Stampa** fornisce le impostazioni per il controllo degli output della tabella per la regressione meta-analisi.

Test del coefficiente del modello
Controlla il test del coefficiente di modello. L'impostazione non è specificata per impostazione predefinita, che sopprime il test. Quando viene specificata l'impostazione, la prova viene fornita nell'output.

Visualizza statistiche con esponenti
Controlla le stime dei parametri. L'impostazione non è specificata per impostazione predefinita, che sopprime le stime dei parametri. Quando viene specificata l'impostazione, le stime dei parametri vengono fornite nell'output.

Definizione Impostazioni di stampa di regressione meta - analisi

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Meta Regressione**
2. Nella finestra di dialogo **Meta - Analysis Regression** , fare clic su **Stampa**.
3. Selezionare e definire le impostazioni di stampa appropriate.
4. Fare clic su **Continua**.

Regressione meta - analisi: Salva

La finestra di dialogo **Salva** fornisce le impostazioni per il salvataggio delle statistiche stimate al dataset attivo per la regressione meta-analisi.

Studi individuali
Fornisce opzioni per il salvataggio della dimensione dell'effetto stimato.

Valori previsti
Predispone e salva i valori montati.

Errore standard dei valori previsti
Prevede e salva l'errore standard dei valori montati.

Limite inferiore intervallo di confidenza
Salva l'intervallo di confidenza stimato inferiore dei valori montati.

Limite superiore intervallo di confidenza
Salva l'intervallo di confidenza stimato per il limite superiore dei valori montati.

Residui
Previene e salva i residui.

Errore standard dei residui
Previene e salva l'errore standard dei residui.

Punti di leva
Predica e salva i leveraggi.

Statistiche cumulative
Fornisce opzioni per il salvataggio delle previsioni lineari.

Nota: Le impostazioni di statistiche cumulative vengono ignorate quando **Fixed - effects** viene selezionato come impostazione **Modello** .

Previsioni lineari fisse
Prevede e salva la previsione lineare.

Errore standard delle previsioni lineari fisse
Previene e salva l'errore standard della previsione lineare.

Previsioni lineari senza distorsione empiriche
Prevede e salva la migliore previsione lineare non distorta degli effetti casuali.

Errore standard dei BLUP
Prevede e salva l'errore standard della migliore previsione lineare non distorta degli effetti casuali.

Definizione Meta - Analisi Regression salva impostazioni

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Meta Regressione**
2. Nella finestra di dialogo **Meta - Analysis Regression** , fare clic su **Salva**.
3. Selezionare e definire le impostazioni di salvataggio appropriate.
4. Fare clic su **Continua**.

Meta - Analisi Regression: Plot

La finestra di dialogo **Plot** fornisce le impostazioni per la definizione dei tracciati delle bolle. Per ulteriori informazioni, consultare “Grafico a bolle” a pagina 79.

Opzioni di tracciamento meta - analisi

Le seguenti opzioni di tracciamento sono disponibili per le procedure di analisi Meta e sono accessibili nelle finestre di dialogo **Plot** .

Grafico Forest

La scheda **Forest Plot** fornisce le impostazioni per il controllo dei grafici della tracciatura delle foreste che visualizzano nell'output per la meta - analisi con esiti continui e binari quando nel dataset attivo vengono forniti dati grezza e di effetto pre - calcolato.

Grafico Forest
L'impostazione opzionale abilita e disabilita l'output della trama forestale. L'impostazione è disabilitata per impostazione predefinita.

Colonne di visualizzazione
Le impostazioni facoltative specificano le statistiche stimate che vengono tabulate.

Dimensione degli effetti
Include la dimensione dell'effetto stimato.

Errore standard
Include l'errore standard della dimensione dell'effetto stimato.

Limiti intervallo di confidenza
Include l'intervallo di confidenza inferiore e i limiti superiori della dimensione dell'effetto stimato.

Valore P
Include il valore *p*della dimensione dell'effetto stimato per ogni singolo studio.

Peso
Include il peso degli studi primari.

Visualizza formato con esponenti

Nota: Questa impostazione si applica solo alla meta - analisi con esiti binari.

Controlla la colonna di trama e la visualizzazione delle dimensioni dell'effetto e dell'intervallo di confidenza. Quando l'impostazione viene selezionata, compi il rapporto di probabilità, il rapporto di probabilità di Peto, o il rapporto di rischio, e tabulare le statistiche in formato esponenziale. Quando l'impostazione non viene selezionata, compendi il rapporto di odds ratio, il rapporto di probabilità di log di Peto o il rapporto di rischio di registrazione e tabulare le statistiche di log - trasformate.

Variabili
Elenca le variabili del dataset disponibili.

Ulteriori colonne
Facoltativamente, selezionare le variabili da visualizzare come colonne aggiuntive. L'ordine variabile determina l'ordine di visualizzazione della colonna.

Ordina
Facoltativamente, specificare una variabile su cui è ordinata la trama forestale. Quando viene specificata una variabile, la trama della foresta è ordinata in ordine **Ascending** (impostazione predefinita). Quando **Descending** viene selezionato, la trama forestale è ordinata in ordine decrescente.

Posizione della colonna del grafico
Controlla il posizionamento della colonna di trama.

Destra
L'impostazione predefinita pone la colonna di trama a destra delle altre colonne della tabella.

A sinistra
Posisce la colonna di trama a sinistra delle altre colonne della tabella.

Righe di riferimento
Controlla le linee di riferimento che si aggiungono alla trama forestale. **Dimensione totale dell'effetto** aggiunge la linea per denotare la dimensione di effetto complessivo stimata. **Dimensione effetto nullo** aggiunge la linea per denotare la dimensione dell'effetto nullo.

Annotazioni
Controlla la visualizzazione delle annotazioni.

Omogeneità
Stampa le statistiche di prova di omogeneità.

Eterogeneità
Stampa le statistiche di prova eterogeneità.

Verifica
Stampa la prova della dimensione di effetto complessivo e la prova di omogeneità tra sottogruppi in presenza dell'analisi del sottogruppo.

Ritaglia intervallo
Specifica la gamma di colture. Quando l'impostazione viene selezionata, devono essere specificati due valori numerici e soddisfare `Upper bound > Lower bound`.

Per la meta - analisi con esiti binari, quando l'impostazione **Effect Size** è impostata su **Differenza di rischio**, deve anche soddisfare `Upper bound ≤ 1` e `Lower bound ≥ −1`. Quando l'impostazione **Visualizza modulo esponenziato** viene selezionata, deve anche soddisfare `Lower bound ≥ 0`.

Definizione delle impostazioni della traccia forestale

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui o binari** > **Dati raw ...** o **Dimensione Effetto Precalcolato ...**
2. Nella finestra di dialogo **Meta - Analisi** , fare clic su **Plot** e quindi fare clic sulla scheda **Plot forestale** .
3. Selezionare e definire le impostazioni appropriate della trama forestale.
4. Fare clic su **Continua**.

Grafico Forest cumulativo

La Scheda **Plot Forest Plot** fornisce le impostazioni per il controllo dei grafici di tracciamento delle foreste cumulabili che visualizzano nell'output per la meta - analisi con esiti continui e binari quando nel dataset attivo vengono forniti dati grezzi e dati di grandezza pre - calcolati.

Grafico Forest cumulativo
L'impostazione opzionale abilita e disabilita l'output della trama forestale. L'impostazione è disabilitata per impostazione predefinita.

Colonne di visualizzazione
Le impostazioni facoltative specificano le statistiche stimate che vengono tabulate.

Dimensione effetto cumulativa
Include la dimensione complessiva dell'effetto cumulativo.

Errore standard
Include l'errore standard della dimensione complessiva dell'effetto cumulativo.

Limiti intervallo di confidenza
Include l'intervallo di confidenza inferiore e i limiti superiori della dimensione complessiva dell'effetto cumulativo.

Valore P
Include il valore *p*della dimensione di effetto cumulativo cumulativo per ogni singolo studio.

Visualizza formato con esponenti

Nota: Questa impostazione si applica solo alla meta - analisi con esiti binari.

Controlla la colonna di trama e la visualizzazione delle dimensioni dell'effetto e dell'intervallo di confidenza. Quando l'impostazione viene selezionata, compone il rapporto di probabilità, il rapporto di probabilità di Peto, o il rapporto di rischio, e tabulare le statistiche in formato esponenziale. Quando l'impostazione non è abilitata, compensa il rapporto di odds ratio, il rapporto di probabilità di log di Peto o il rapporto di rischio di registrazione e tabulare le statistiche di registrazione log.

Variabili
Elenca le variabili del dataset disponibili.

Ulteriori colonne
Facoltativamente, selezionare le variabili da visualizzare come colonne aggiuntive. L'ordine variabile determina l'ordine di visualizzazione della colonna.

Posizione della colonna del grafico
Controlla il posizionamento della colonna di trama.

Destra
L'impostazione predefinita pone la colonna di trama a destra delle altre colonne della tabella.

A sinistra
Posisce la colonna di trama a sinistra delle altre colonne della tabella.

Ritaglia intervallo
Specifica la gamma di colture. Quando l'impostazione viene selezionata, devono essere specificati due valori numerici e soddisfare `Upper bound > Lower bound`.

Per la meta - analisi con esiti binari, quando l'impostazione **Effect Size** è impostata su **Differenza di rischio**, deve anche soddisfare `Upper bound ≤ 1` e `Lower bound ≥ −1`. Quando l'impostazione **Visualizza modulo esponenziato** viene selezionata, deve anche soddisfare `Lower bound ≥ 0`.

Definizione delle impostazioni di plot forest cumulative

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui o binari** > **Dati raw ...** o **Dimensione Effetto Precalcolato ...**
2. Nella finestra di dialogo **Meta - Analisi** , fare clic su **Plot** e quindi fare clic sulla scheda **Cumulative Forest Plot** .
3. Selezionare e definire le impostazioni di tracciamento delle foreste cumulative appropriate.
4. Fare clic su **Continua**.

Grafico a bolle

La Scheda **Bubble Plot** fornisce le impostazioni per il controllo dei grafici della tracciatura delle bolle che visualizzano nell'output per la meta - analisi con esiti continui e binari quando nel dataset attivo vengono forniti dati grezzi e dati di grandezza pre - calcolati.

Nota: Bubble Plot è l'unico tipo di trama disponibile per l'analisi di regressione meta-analisi.

Grafico a bolle
L'impostazione opzionale abilita e disabilita l'output della trama bolla. L'impostazione è disabilitata per impostazione predefinita.

Nota: Questa impostazione non è disponibile per l'analisi di regressione meta-analisi.

Variabili
Elenca le variabili del dataset disponibili.

Predittore (s)
Specificare le variabili trattate come predittori continui sull'asse x. Sono consentite più variabili con ciascuna creazione di una trama a bolla separata.

Nota: Per l'analisi di regressione meta - analisi, questa impostazione è etichettata **Bubble Plot Predictor (s)** e consente di selezionare covariate per le quali creare grafici a bolle.

Predittori centrati sulla media
L'impostazione opzionale controlla il predittore continuo sull'asse x. L'impostazione è disabilitata per impostazione predefinita, che non centra il predittore. Quando l'impostazione è abilitata, il predittore è mediamente centrato.

Etichetta
Facoltativamente specificare una variabile su cui la trama a bolla è etichettata. L'elenco **Posizionamento** fornisce le opzioni per collocare le etichette automaticamente (**Auto**) o sulle **Giusto**, **Sinistro**, **Superiore**e **Fondo**.

Disegna bolle proporzionali ai pesi
Controlla come sono tracciate le bolle in relazione ai pesi. L'impostazione è abilitata per impostazione predefinita, che richiama le bolle proporzionali ai pesi. Quando l'impostazione è disabilitata, tutte le bolle sono disegnate alla stessa dimensione.

Visualizza la linea adattata
Controlla la visualizzazione della linea di regressione adattata. L'impostazione è abilitata per impostazione predefinita, che aggiunge la linea di regressione adattata alla trama della bolla. Quando l'impostazione è disabilitata, la linea di regressione adattata non viene aggiunta alla trama della bolla.

Visualizza i limiti dell'intervallo di confidenza
Controlla la visualizzazione dei limiti dell'intervallo di confidenza. L'impostazione è abilitata per impostazione predefinita, che aggiunge l'intervallo di confidenza limiti alla trama della bolla. Quando l'impostazione è disabilitata, i limiti dell'intervallo di confidenza non vengono aggiunti alla trama della bolla.

Intervallo dell'asse X
Specifica la gamma di trama dell'asse x. Quando l'impostazione è abilitata, devono essere specificati due valori numerici e soddisfare `Upper bound > Lower bound`. Da notare che i valori specificati si applicano a tutti i tracciati delle bolle.

Intervallo dell'asse Y
Specifica la gamma di trama dell'asse y. Quando l'impostazione è abilitata, devono essere specificati due valori numerici e soddisfare `Upper bound > Lower bound`. Da notare che i valori specificati si applicano a tutti i tracciati delle bolle.

Per la meta - analisi con esiti binari, quando l'impostazione **Effect Size** è impostata su **Differenza di rischio**, deve anche soddisfare `Upper bound ≤ 1` e `Lower bound ≥ −1`.

Definizione delle impostazioni della trama

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui o binari** > **Dati raw ...** o **Dimensione Effetto Precalcolato ...**

   oppure

   Analizzare > **Meta Analisi** > **Meta Regressione**
2. Nella finestra di dialogo **Meta - Analisi** , fare clic su **Plot** e quindi fare clic sulla scheda **Bubble Plot** .

3. Selezionare e definire le impostazioni di tracciamento delle bolle appropriate.
4. Fare clic su **Continua**.

Grafico a imbuto

La scheda **Funnel Plot** fornisce le impostazioni per il controllo dei grafici della trama del funnel che visualizzano nell'output per la meta - analisi con esiti continui e binari quando nel dataset attivo vengono forniti dati grezza e di effetto pre - calcolato.

Grafico a imbuto
L'impostazione opzionale abilita e disabilita l'output della trama del funnel. L'impostazione è disabilitata per impostazione predefinita.

Valori asse Y
Controlla i valori dell'asse y.

Errore standard
Compi l'errore standard.

Errore standard inverso
Compi l'errore standard inverso.

Varianza
Compensa la varianza.

Varianza inversa
Compensa la varianza inversa.

Intervallo dell'asse X
Specifica la gamma di trama dell'asse x. Quando l'impostazione è abilitata, devono essere specificati due valori numerici e soddisfare `Upper bound > Lower bound`. Da notare che i valori specificati si applicano a tutti i tracciati del funi.

Per la meta - analisi con esiti binari, quando l'impostazione **Effect Size** è impostata su **Differenza di rischio**, deve anche soddisfare `Upper bound ≤ 1` e `Lower bound ≥ −1`.

Intervallo dell'asse Y
Specifica la gamma di trama dell'asse y. Quando l'impostazione è abilitata, devono essere specificati due valori numerici e soddisfare `Upper bound > Lower bound ≥ 0`. Da notare che i valori specificati si applicano a tutti i tracciati del funi.

Includi studi assegnati con ritaglia e riempi
Controlla se gli studi imputati sono inclusi nella trama del funnel quando l'impostazione **Stima numero di studi mancanti** è abilitata nella finestra di dialogo **Trim - and - Fill** . Le impostazioni sono disabilitate per impostazione predefinita, che esclude gli studi imputati. Quando **Visualizza la dimensione complessiva degli effetti degli studi osservati** viene selezionata, si aggiunge una linea di riferimento verticale che denota la dimensione di effetto complessivo stimata. L'impostazione tiene conto sia degli studi osservati che di quelli imputati.

Variabili
Elenca le variabili del dataset disponibili.

Etichetta
Facoltativamente specificare una variabile su cui è etichettata la trama del funnel. L'elenco **Posizionamento** fornisce le opzioni per collocare le etichette automaticamente (**Auto**) o sulle **Giusto**, **Sinistro**, **Superiore**e **Fondo**.

Definizione delle impostazioni del tracciato del funnel

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui o binari** > **Dati raw ...** o **Dimensione Effetto Precalcolato ...**
2. Nella finestra di dialogo **Meta - Analisi** , fare clic su **Plot** e quindi fare clic sulla scheda **Plot Funnel** .
3. Selezionare e definire le impostazioni del tracciato del funnel appropriato.

4. Fare clic su **Continua**.

Grafico Galbraith

La scheda **Galbraith Plot** fornisce le impostazioni per il controllo dei grafici della trama di Galbraith che visualizzano nell'output per la meta - analisi con esiti continui e binari quando nel dataset attivo vengono forniti dati grezze e dati di grandezza pre - calcolati.

Grafico Galbraith
L'impostazione opzionale abilita e disabilita l'output della trama bolla. L'impostazione è disabilitata per impostazione predefinita.

Visualizza i limiti dell'intervallo di confidenza
Controlla la visualizzazione dei limiti dell'intervallo di confidenza. L'impostazione è abilitata per impostazione predefinita, che aggiunge l'intervallo di confidenza limiti alla trama della bolla. Quando l'impostazione è disabilitata, i limiti dell'intervallo di confidenza non vengono aggiunti alla trama della bolla.

Intervallo dell'asse X
Specifica la gamma di trama dell'asse x. Quando l'impostazione è abilitata, devono essere specificati due valori numerici e soddisfare `Upper bound > Lower bound ≥ 0`. Da notare che i valori specificati si applicano a tutti i tracciati di Galbraith.

Intervallo dell'asse Y
Specifica la gamma di trama dell'asse y. Quando l'impostazione è abilitata, devono essere specificati due valori numerici e soddisfare `Upper bound > Lower bound`. Da notare che i valori specificati si applicano a tutti i tracciati di Galbraith.

Variabili
Elenca le variabili del dataset disponibili.

Etichetta
Facoltativamente specificare una variabile su cui la trama a bolla è etichettata. L'elenco **Posizionamento** fornisce le opzioni per collocare le etichette automaticamente (**Auto**) o sulle **Giusto**, **Sinistro**, **Superiore**e **Fondo**.

Definizione delle impostazioni della trama Galbraith

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui o binari** > **Dati raw ...** o **Dimensione Effetto Precalcolato ...**
2. Nella finestra di dialogo **Meta - Analisi** , fare clic su **Plot** e quindi fare clic sulla scheda **Plot Galbraith** .
3. Selezionare e definire le impostazioni della trama Galbraith appropriate.
4. Fare clic su **Continua**.

L'Abb'e Plot

La scheda **L'Abb'e Plot** fornisce le impostazioni per il controllo dei grafici della trama di L'Abb'e che visualizzano nell'output per la meta - analisi con esiti continui e binari quando nel dataset attivo vengono forniti dati grezzi e dati di grandezza pre - calcolati.

La trama di Abb'e
L'impostazione opzionale abilita e disabilita l'output della trama L'Abb'e. L'impostazione è disabilitata per impostazione predefinita.

Disegna punti proporzionali ai pesi
Controlla come sono tracciati i doti in relazione ai pesi. L'impostazione è abilitata per impostazione predefinita, che richiama i doti proporzionali ai pesi. Quando l'impostazione è disabilitata, tutti i puntini sono disegnati alla stessa dimensione.

Righe di riferimento

Dimensione effetto complessiva
Controlla la visualizzazione della linea di riferimento che rappresenta la dimensione di effetto complessivo stimata. L'impostazione è disabilitata per impostazione predefinita, che sopprime la linea di riferimento. La riga di riferimento visualizza quando l'impostazione è abilitata.

Dimensione effetto nulla
Controlla la visualizzazione della linea di riferimento che non rappresenta alcun effetto. L'impostazione è abilitata per impostazione predefinita, che visualizza la linea di riferimento. La linea di riferimento viene soppressa quando l'impostazione non è abilitata.

Intervallo dell'asse X
Specifica la gamma di trama dell'asse x. Quando l'impostazione è abilitata, devono essere specificati due valori numerici e soddisfare `Upper bound > Lower bound`. Da notare che i valori specificati si applicano a tutti i tracciati di Galbraith.

Per la meta - analisi con esiti binari, quando l'impostazione **Effect Size** è impostata su **Differenza di rischio**, deve anche soddisfare `Lower bound ≥ 0` e `Upper bound ≤ 1`. Quando l'impostazione **Effect Size** è impostata su **Log Risk Ratio**, deve soddisfare `Upper bound ≤ 0`.

Intervallo dell'asse Y
Specifica la gamma di trama dell'asse y. Quando l'impostazione è abilitata, devono essere specificati due valori numerici e soddisfare `Upper bound > Lower bound`. Da notare che i valori specificati si applicano a tutti i tracciati di Galbraith.

Per la meta - analisi con esiti binari, quando l'impostazione **Effect Size** è impostata su **Differenza di rischio**, deve anche soddisfare `Lower bound ≥ 0` e `Upper bound ≤ 1`. Quando l'impostazione **Effect Size** è impostata su **Log Risk Ratio**, deve soddisfare `Upper bound ≤ 0`.

Variabili
Elenca le variabili del dataset disponibili.

Etichetta
Facoltativamente specificare una variabile su cui la trama a bolla è etichettata. L'elenco **Posizionamento** fornisce le opzioni per collocare le etichette automaticamente (**Auto**) o sulle **Giusto**, **Sinistro**, **Superiore**e **Fondo**.

Definizione delle impostazioni della trama di 'Abb'e

1. Dai menu, scegliere:

   Analizzare > **Meta Analisi** > **Esiti continui o binari** > **Dati raw ...**
2. Nella finestra di dialogo **Meta - Analisi** , fare clic su **Plot** e quindi fare clic sulla scheda **L'Abb'e Plot** .
3. Selezionare e definire le impostazioni della trama dell'Abb'e appropriata.
4. Fare clic su **Continua**.

Informazioni sui dati

Le informazioni sui dati restituiscono informazioni del dizionario, ad esempio nomi di variabili, etichette di variabile e valore o valori mancanti, e statistiche di riepilogo per alcune o tutte le variabili e gli insiemi a risposta multipla presenti nel dataset attivo. Per le variabili nominali e ordinali e gli insiemi a risposta multipla, le statistiche di riepilogo includono conteggi e percentuali. Per le variabili di scala, le statistiche di riepilogo includono media, deviazione standard e quartili.

Nota: le informazioni sui dati ignorano lo stato del file suddiviso. Sono inclusi i gruppi di file suddivisi creati per i valori mancanti valori minimo, massimo e di incremento (disponibili nel modulo aggiuntivo Missing Values).

Per ottenere le informazioni sui dati

1. Dai menu, scegliere:

   Analizzare > **Report** > **Codebook**

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Fare clic sulla scheda Variabili.
3. Selezionare una o più variabili e/o uno o più insiemi a risposta multipla.

Facoltativamente, è possibile:

- Controllare le informazioni sulle variabili visualizzate.
- Controllare le statistiche visualizzate (o escludere tutte le statistiche di riepilogo).
- Controllare l'ordine in cui vengono visualizzati insiemi a risposta multipla e variabili.
- Modificare il livello di misurazione per le variabili nell'elenco di origine in modo tale da modificare le statistiche di riepilogo visualizzate. Per ulteriori informazioni, consultare l'argomento "Scheda Informazioni sui dati - Statistiche" a pagina 86.

Modifica del livello di misurazione

È possibile modificare temporaneamente il livello di misurazione per le variabili. Non è possibile modificare il livello di misurazione per gli insiemi a risposta multipla, che vengono sempre trattati come nominali)

1. Fare clic con il pulsante destro del mouse su una variabile nell'elenco origine.
2. Selezionare un livello di misurazione dal menu a comparsa.

In questo modo il livello di misurazione viene temporaneamente modificato. Da un punto di vista pratico, è utile solo per le variabili numeriche. Il livello di misurazione per le variabili stringa è limitato a nominale o ordinale, entrambi trattati allo stesso modo dalla procedura Informazioni sui dati.

Scheda Output della finestra Informazioni sui dati

La scheda Output controlla le informazioni sulle variabili incluse per ciascuna variabile e ciascun insieme a risposta multipla, l'ordine in cui variabili e insiemi a risposta multipla vengono visualizzati e il contenuto della tabella delle informazioni facoltative sui file.

Informazioni sulla variabile

Controlla le informazioni del dizionario visualizzate per ciascuna variabile.

Posizione. Intero che rappresenta la posizione della variabile nell'ordine del file. Non è disponibile per gli insiemi a risposta multipla.

Etichetta. Etichetta descrittiva associata alla variabile o all'insieme a risposta multipla.

Tipo. Tipo di dati fondamentale. Può essere *Numerico*, *Stringa* o *Insieme a risposta multipla*.

Formato. Formato di visualizzazione della variabile, ad esempio *A4*, *F8.2* o *DATE11*. Non è disponibile per gli insiemi a risposta multipla.

Livello di misurazione. I valori possibili sono *Nominale*, *Ordinale*, *Scala* e *Sconosciuto*. Il valore visualizzato è il livello di misurazione memorizzato nel dizionario e non è influenzato da alcuna variazione temporanea del livello di misurazione dovuta alla modifica del livello nell'elenco delle variabili di origine della scheda Variabili. Non è disponibile per gli insiemi a risposta multipla.

Nota: il livello di misurazione per le variabili numeriche può essere "sconosciuto" prima che vengano forniti i primi dati quando il livello di misurazione non è stato impostato in modo esplicito, ad esempio, i dati letti da un'origine esterna o le variabili appena create. Per ulteriori informazioni, consultare l'argomento .

Ruolo. Alcune finestre di dialogo supportano la capacità di pre-selezionare le variabili per l'analisi in base a dei ruoli definiti.

Etichette valori. Etichette descrittive associate a valori dei dati specifici.

- Se nella scheda Statistiche è selezionata l'opzione Conteggio o Percentuale, le etichette valore definiti sono incluse nell'output anche se l'opzione Etichette valore non è stata selezionata.

- Per gli insiemi a dicotomie multiple, le "etichette valore" sono le etichette di variabile per le variabili elementari presenti nell'insieme o le etichette dei valori conteggiati, in base alla definizione dell'insieme. Per ulteriori informazioni, consultare l'argomento .

Valori mancanti. Valori mancanti definiti dall'utente. Se si seleziona Conteggio o Percentuale nella scheda Statistiche, le etichette valore definite vengono incluse nell'output anche se non si seleziona Valori mancanti qui. Non è disponibile per gli insiemi a risposta multipla.

Attributi personalizzati. Attributi della variabile definiti dall'utente. L'output include sia i nomi sia i valori di tutti gli attributi della variabile personalizzati associati a ciascuna variabile. Per ulteriori informazioni, consultare l'argomento . Non è disponibile per gli insiemi a risposta multipla.

Attributi riservati. Attributi della variabile di sistema riservati. È possibile visualizzare gli attributi di sistema, ma non modificarli. I nomi degli attributi di sistema iniziano con un segno di dollaro ($). Gli attributi non visualizzati, con nomi che iniziano con "@" o con "$@", non sono inclusi. L'output include sia i nomi sia i valori di tutti gli attributi di sistema associati a ciascuna variabile. Non è disponibile per gli insiemi a risposta multipla.

Informazioni sul file

La tabella delle informazioni opzionali sul file può includere i seguenti attributi:

Nome file. Nome del file di dati di IBM SPSS Statistica . Se il dataset non è mai stato salvato in formato IBM SPSS Statistica , non esiste alcun nome di file di dati. Se non è visualizzato un nome di file nella barra del titolo della finestra Editor dei dati, il dataset attivo non ha un nome di file.

Ubicazione. Ubicazione della cartella del file di dati di IBM SPSS Statistica . Se il dataset non è mai stato salvato in formato IBM SPSS Statistica , non esiste alcuna ubicazione.

Numero di casi. Numero di casi presenti nel dataset attivo. Si tratta del numero totale di casi, compresi gli eventuali casi esclusi dalle statistiche di riepilogo a causa delle condizioni di filtro.

Etichetta. Questa è l'etichetta del file (se presente) definita dal comando `FILE LABEL` .

Documenti. Testo del documento del file di dati.

Stato del peso. Se il peso è attivo, viene visualizzato il nome della variabile peso. Per ulteriori informazioni, consultare l'argomento .

Attributi personalizzati. Attributi del file di dati definiti dall'utente. Gli attributi del file di dati vengono definiti con il comando `DATAFILE ATTRIBUTE`.

Attributi riservati. Attributi riservati del file di dati di sistema. È possibile visualizzare gli attributi di sistema, ma non modificarli. I nomi degli attributi di sistema iniziano con un segno di dollaro ($). Gli attributi non visualizzati, con nomi che iniziano con "@" o con "$@", non sono inclusi. L'output include sia i nomi che i valori di tutti gli attributi del file di dati di sistema.

Ordine di visualizzazione variabili

Sono disponibili le seguenti opzioni per il controllo dell'ordine in cui vengono visualizzati gli insiemi a risposta multipla e le variabili.

Alfabetico. Ordine alfabetico per nome di variabile.

File. Ordine in cui le variabili appaiono nel dataset (l'ordine in cui sono visualizzate nell'Editor dei dati). In ordine crescente, con gli insiemi a risposta multipla visualizzati per ultimi, dopo tutte le variabili selezionate.

Livello di misurazione. Ordina per livello di misurazione. Crea quattro gruppi di ordinamento: nominale, ordinale, scala e sconosciuto. Gli insiemi a risposta multipla vengono trattati come nominali.

Nota: il livello di misurazione per le variabili numeriche può essere "sconosciuto" prima che vengano forniti i primi dati quando il livello di misurazione non è stato impostato in modo esplicito, ad esempio, i dati letti da un'origine esterna o le variabili appena create.

Elenco di variabili. Ordine in cui le variabili e gli insiemi a risposta multipla appaiono nell'elenco delle variabili selezionate della scheda Variabili.

Nome attributo personalizzato. L'elenco delle opzioni del criterio di ordinamento include anche i nomi degli attributi della variabile personalizzati definiti dall'utente. In ordine crescente, con le variabili senza attributi per prime, seguite dalle variabili che hanno un attributo ma senza valori definiti per lo stesso, seguite dalle variabili con valori definiti per l'attributo e valori in ordine alfabetico.

Numero massimo di categorie

Se l'output include etichette valore, conteggi o percentuali per ogni valore univoco, è possibile eliminare queste informazioni dalla tabella se il numero di valori supera il valore specificato. Per impostazione predefinita, queste informazioni non vengono inserite se il numero di valori univoci per la variabile supera 200.

Scheda Informazioni sui dati - Statistiche

La scheda Statistiche consente di controllare le statistiche di riepilogo incluse nell'output o di eliminare completamente la visualizzazione delle statistiche di riepilogo.

Conteggi e percentuali

Per le variabili nominali e ordinali, gli insiemi a risposta multipla e i valori con etichetta delle variabili di scala, sono disponibili le statistiche riportate di seguito.

Conteggio. Il conteggio o numero di casi con ogni valore (o intervallo di valori) di una variabile.

Percentuale. La percentuale di casi con un valore particolare.

Tendenza centrale e dispersione

Per le variabili di scala, le statistiche disponibili sono:

Media. Una misura della tendenza centrale. La media aritmetica, ossia la somma divisa per il numero di casi.

deviazione standard. Misura della dispersione intorno alla media. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95% dei casi rientra in due deviazioni standard. Se, ad esempio, l'età media è 45 e la deviazione standard è 10, il 95% dei casi sarà compreso tra 25 e 65 in un distribuzione normale.

Quartili. Mostra i valori corrispondenti al 25°, 50° e 75° percentile.

Nota: è possibile modificare temporaneamente il livello di misurazione associato a una variabile (e, quindi, modificare le statistiche di riepilogo visualizzate per tale variabile) nell'elenco di variabili di origine nella scheda Variabili.

Frequenze

La procedura Frequenze consente di ottenere statistiche e rappresentazioni grafiche che risultano utili per la descrizione di molti tipi di variabili. La procedura Frequenza offre un'ottima opportunità per iniziare ad osservare i dati.

Per ottenere un report e un grafico a barre delle frequenze è possibile disporre i singoli valori in ordine crescente o decrescente oppure ordinare le categorie in base alle rispettive frequenze. Il report sulle frequenze può essere eliminato se una variabile ha molti valori distinti. È possibile etichettare i grafici con frequenze (valore predefinito) o percentuali.

Esempio
Qual è la distribuzione dei clienti di un'azienda per tipo di industria? Dall'output, si nota che il 37,5% dei clienti fa parte di enti governativi, il 24,9% fa parte di società, il 28,1% di istituzioni accademiche e il 9,4% del settore sanitario. Per i dati quantitativi e continui, ad esempio il fatturato, si può notare che la vendita media del prodotto è pari a €. 3.576 con una deviazione standard di €. 1.078.

Statistiche e grafici
Conteggi della frequenza, percentuali, percentuali cumulative, media, mediana, moda, somma, deviazione standard, varianza, intervallo, valori minimo e massimo, errore standard della media,

asimmetria e curtosi (entrambe con errori standard), quartili, percentili definiti dall'utente, grafici a barre, grafici a torta e istogrammi.

Considerazioni sui dati

Dati
Utilizzare codici numerici o stringhe per codificare le variabili categoriali (misure di livello nominale o ordinale).

Ipotesi
Le tabulazioni e le percentuali forniscono un'utile descrizione dei dati provenienti da qualsiasi distribuzione, in particolare per le variabili con categorie ordinate o non ordinate. La maggior parte delle statistiche di riepilogo, ad esempio la media e la deviazione standard, si basano sulla normale teoria e sono idonee per variabili quantitative con distribuzioni simmetriche. Le statistiche robuste, ad esempio la media, i quartili e i percentili, sono idonee per variabili quantitative rispondenti o meno all'ipotesi di normalità.

Ottenere tabelle di frequenza

1. Dai menu, scegliere:

   Analizza > **Statistiche descrittive** > **Frequenze...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili categoriali o quantitative.
3. Opzionalmente, selezionare **Crea tabelle di stile APA** per creare tabelle di output che aderiscono alle linee guida in stile APA.
4. Facoltativamente, è possibile:
   - Fare clic su **Statistiche** per ottenere statistiche descrittive per le variabili quantitative.
   - Fare clic su **Grafici** per ottenere grafici a barre, grafici a torta e istogrammi.
   - Fare clic su **Formato** per stabilire l'ordine in cui visualizzare i risultati.
   - Clicca su **Stile** per specificare le condizioni per la modifica automatica delle proprietà delle tabelle pivot in base a specifiche condizioni.
   - Clicca su **Bootstrap** per ricavare stime robusta degli errori standard e degli intervalli di confidenza per le stime come la media, la mediana, la proporzione, il rapporto di probabilità, il coefficiente di correlazione o il coefficiente di regressione. Può essere utilizzato anche per creare dei test sull'ipotesi.

Frequenze: Statistiche

Valori percentili. Valori di una variabile quantitativa che suddividono i dati ordinati in due gruppi in modo da visualizzare una percentuale sopra e una sotto. I quartili (il 25°, 50° e 75° percentile) suddividono le osservazioni in quattro gruppi di dimensioni uguali. Se si desidera avere un numero uguale di gruppi diverso da quattro, selezionare **Punti di divisione per: n gruppi uguali**. È inoltre possibile specificare i singoli percentili, ad esempio il 95° percentile, ovvero il valore al di sotto del quale ricade il 95% delle osservazioni.

Tendenza centrale. Le statistiche che descrivono l'ubicazione della distribuzione includono media, mediana, moda e somma di tutti i valori.

- *Media*. Una misura della tendenza centrale. La media aritmetica, ossia la somma divisa per il numero di casi.
- *Median*. Il valore al di sopra e al di sotto del quale ricade la metà dei casi: il 50° percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori estremamente bassi o alti.

- *Modalità*. Il valore che ricorre più frequentemente. Se più valori condividono la maggiore ricorrenza, ognuno di essi è una moda. La procedura Frequenze riporta solo la più piccola di queste mode multiple.
- *Somma*. La somma o il totale di tutti i valori non mancanti di tutti i casi.

Dispersione. Le statistiche che misurano l'entità della variazione o della diffusione dei dati includono deviazione standard, varianza, intervallo, valore minimo e massimo ed errore standard della media.

- *Dev deviazione*. Misura della dispersione intorno alla media. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95% dei casi rientra in due deviazioni standard. Se, ad esempio, l'età media è 45 e la deviazione standard è 10, il 95% dei casi sarà compreso tra 25 e 65 in un distribuzione normale.
- *Varianza*. Una misura di dispersione intorno alla media, uguale alla somma dei quadrati delle deviazioni dalla media divisa per un valore inferiore al numero di casi. La varianza è espressa in quadrati dell'unità di misura della variabile.
- *Intervallo*. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.
- *Minimo*. Il valore più basso di una variabile numerica.
- *maximum*. Il valore più alto di una variabile numerica.
- *S. E. media*. Una misura di quanto il valore della media può variare da campione a campione preso dalla stessa distribuzione. Può essere usata per comparare genericamente la media osservata rispetto a un valore ipotizzato (ovvero, è possibile concludere che i due valori sono diversi se il rapporto della differenza rispetto all'errore standard è inferiore a -2 o maggiore di +2).
- *Curriculum vitae*. Il rapporto fra la deviazione standard e la media. I valori sono espressi in percentuale.

Distribuzione. L'asimmetria e la curtosi sono statistiche che descrivono la forma e la simmetria della distribuzione. Queste statistiche vengono visualizzate con i relativi errori standard.

- *Asimmetria*. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con asimmetria positiva ha una coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica uno scostamento dalla simmetria.
- *Curtosi*. Una misura della misura in cui esistono valori anomali. Per la distribuzione normale, il valore della statistica di curtosi è zero. Una curtosi positiva indica che i dati mostrano valori anomali più estremi di una distribuzione normale. La kurtosis negativa indica che i dati esibiano valori anomali meno estremi di una distribuzione normale. La definizione di kurtosi che si usa, dove il valore è di 0 per una distribuzione normale, è a volte indicata come kurtosi in eccesso. Alcuni software possono segnalare kurtosi tale che il valore è di 3 per una distribuzione normale.

I valori sono punti centrali di gruppi. Se i valori dei dati sono punti centrali di gruppi (ad esempio, l'età delle persone sulla trentina è codificata come 35), selezionare questa opzione per valutare la media e i percentili per i dati originali non raggruppati.

Frequenze: Grafici

Nota: I grafici non vengono prodotti nell'output quando **Perform bootstrap** è abilitato sulla finestra di dialogo Bootstrap.

Tipo di diagramma

I grafici a torta mostrano il contributo delle parti all'intero grafico. Ogni sezione di un grafico a torta corrisponde a un gruppo definito da una singola variabile di raggruppamento. Nei grafici a barre il conteggio relativo a ciascun valore o categoria viene rappresentato come una barra distinta, in modo da poter confrontare visivamente le categorie. Anche gli istogrammi contengono barre, che però sono tracciate lungo una scala per intervalli uguali. L'altezza di ogni barra rappresenta il conteggio dei valori di una variabile quantitativa che rientra nell'intervallo. Nell'istogramma vengono indicati la forma, il centro e la diffusione della distribuzione. Una curva normale sovrapposta all'istogramma consente di valutare se i dati sono distribuiti normalmente.

Valori del grafico

Per i grafici a barre, l'asse di scala può essere etichettato in base a percentuali o conteggi della frequenza.

Frequenze: Formato

Ordina per. La tabella delle frequenze può essere disposta in base ai valori effettivi dei dati oppure in base al conteggio (frequenza di ricorrenza) di tali valori, in ordine crescente o decrescente. Se, tuttavia, si desidera ottenere un istogramma o i percentili, si presume che la variabile sia quantitativa e i suoi valori vengano visualizzati in ordine crescente.

Variabili multiple. Se si producono delle tabelle di statistiche per più variabili, è possibile visualizzare tutte le variabili in una sola tabella (**Confronta variabili**) o visualizzare una tabella di statistiche separata per ogni variabile (**Organizza l'output per variabili**).

Sopprimi le tabelle con molte categorie. Questa opzione consente di disattivare la visualizzazione delle tabelle che includono un numero di valori maggiore di quello specificato.

Descrittive

La procedura Descrittive consente di visualizzare statistiche di riepilogo univariate per diverse variabili incluse nella stessa tabella e di calcolare i valori standardizzati (punteggi *z*). È possibile ordinare le variabili in base alle dimensioni delle rispettive medie (in ordine crescente o decrescente), in ordine alfabetico oppure nell'ordine in cui sono state selezionate (impostazione predefinita).

I punteggi *z* salvati vengono aggiunti ai dati nell'Editor dei dati e sono disponibili per la creazione di grafici, elenchi di dati e analisi. Quando le variabili vengono registrate in unità diverse (ad esempio, prodotto interno lordo pro capite e percentuale di alfabetizzazione), una trasformazione dei punteggi *z* consente di posizionare le variabili su una scala comune per facilitarne il confronto visivo.

Esempio. Se ciascun caso incluso nei dati contiene i totali delle vendite giornaliere relativi a ciascun agente di vendita (ad esempio, una voce per Roberto, una per Carlo e una per Bruno), registrati ogni giorno per diversi mesi, la procedura Descrittive consente di calcolare la media delle vendite giornaliere per ogni agente e di ordinare i risultati dalla media di vendita maggiore alla minore.

Statistiche. Dimensioni del campione, media, valore minimo e massimo, deviazione standard, varianza, intervallo, somma, errore standard della media, curtosi e asimmetria degli errori standard.

Considerazioni sui dati relative alla procedura Descrittive

Dati. Utilizzare variabili numeriche dopo averle valutate graficamente per registrare errori, valori anomali e anomalie distributive. La procedura Descrittive risulta molto utile quando si utilizzano file di grandi dimensioni (migliaia di casi).

Ipotesi. La maggior parte delle statistiche disponibili (compresi i punteggi *z*) si fondano sulla teoria di normalità e possono essere utilizzate per le variabili quantitative (misurazioni a livello di intervallo o di rapporto) con distribuzioni simmetriche. Evitare variabili con categorie non ordinate o distribuzioni asimmetriche. La distribuzione dei punteggi *z* ha la stessa forma di quella dei dati originali. Pertanto, il calcolo dei punteggi *z* non rappresenta una soluzione per dati problematici.

Per ottenere statistiche descrittive

1. Dai menu, scegliere:

   Analizza > **Statistiche descrittive** > **Descrittive...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili.

Facoltativamente, è possibile:

- Selezionare **Salva valori standardizzati come variabili** per salvare i punteggi *z* come nuove variabili.
- Fare clic su **Opzioni** per ottenere le statistiche e l'ordine di visualizzazione facoltativi.

Descrittive: Opzioni

Media e somma. Per impostazione predefinita, viene visualizzata la media o la media aritmetica.

Dispersione. Le statistiche che misurano la dispersione o la variazione dei dati includono deviazione standard, varianza, intervallo, valore minimo e massimo ed errore standard della media.

- *Dev deviazione*. Misura della dispersione intorno alla media. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95% dei casi rientra in due deviazioni standard. Se, ad esempio, l'età media è 45 e la deviazione standard è 10, il 95% dei casi sarà compreso tra 25 e 65 in un distribuzione normale.
- *Varianza*. Una misura di dispersione intorno alla media, uguale alla somma dei quadrati delle deviazioni dalla media divisa per un valore inferiore al numero di casi. La varianza è espressa in quadrati dell'unità di misura della variabile.
- *Intervallo*. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.
- *Minimo*. Il valore più basso di una variabile numerica.
- *maximum*. Il valore più alto di una variabile numerica.
- *E. S. della media*.Una misura di quanto il valore della media può variare da campione a campione preso dalla stessa distribuzione. Può essere usata per comparare genericamente la media osservata rispetto a un valore ipotizzato (ovvero, è possibile concludere che i due valori sono diversi se il rapporto della differenza rispetto all'errore standard è inferiore a -2 o maggiore di +2).
- *Curriculum vitae*. Il rapporto fra la deviazione standard e la media. I valori sono espressi in percentuale.

Distribuzione. Curtosi e asimmetria sono statistiche che caratterizzano la forma e la simmetria della distribuzione. Queste statistiche vengono visualizzate con i relativi errori standard.

- *Curtosi*. Una misura della misura in cui esistono valori anomali. Per la distribuzione normale, il valore della statistica di curtosi è zero. Una curtosi positiva indica che i dati mostrano valori anomali più estremi di una distribuzione normale. La kurtosis negativa indica che i dati esibiano valori anomali meno estremi di una distribuzione normale. La definizione di kurtosi che si usa, dove il valore è di 0 per una distribuzione normale, è a volte indicata come kurtosi in eccesso. Alcuni software possono segnalare kurtosi tale che il valore è di 3 per una distribuzione normale.
- *Asimmetria*. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con asimmetria positiva ha una coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica uno scostamento dalla simmetria.

Ordine di visualizzazione. Per impostazione predefinita, le variabili vengono visualizzate nell'ordine in cui vengono selezionate. È inoltre possibile visualizzare le variabili in ordine alfabetico, per media crescente o per media decrescente.

Funzioni aggiuntive del comando DESCRIPTIVES

Il linguaggio della sintassi dei comandi consente inoltre di:

- Salvare i punteggi standardizzati (punteggi *z*) per alcune ma non per tutte le variabili (con il sottocomando `VARIABLES`).
- Specificare i nomi delle nuove variabili che contengono i punteggi standardizzati (con il sottocomando `VARIABLES`).
- Escludere dall'analisi i casi con valori mancanti per qualsiasi variabile (con il sottocomando `MISSING`).
- Ordinare le variabili visualizzate in base al valore di una statistica, non solo in base alla media (con il sottocomando `SORT`).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Percentili

La procedura Percentili visualizza tabelle di percentili.

Valori percentili
Quando si seleziona **Valori percentili**, vengono visualizzati per impostazione predefinita i valori relativi al 5°, 10°, 25°, 50°, 75°, 90° e 95° percentile. Valori di una variabile quantitativa che suddividono i dati ordinati in due gruppi in modo da visualizzare una percentuale sopra e una sotto.

Quartili
I quartili (il 25°, 50° e 75° percentile) suddividono le osservazioni in quattro gruppi di dimensioni uguali.

Personalizzato
Quando si seleziona **Personalizzato**, immettere almeno un valore per eseguire l'analisi. I valori di input devono essere un numero compreso tra 0 e 100. Utilizzare i pulsanti **Aggiungi**, **Modifica**e **Rimuovi** per gestire i valori nell'elenco dei valori percentili.

Metodo percentile
Per impostazione predefinita, il metodo HAVERAGE è selezionato per calcolare i percentili.

Come ottenere le tabelle percentili

1. Dai menu, scegliere:

   Analisi > **Statistiche descrittive** > **Percentili ..**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili numeriche e spostarle nel campo Variabili.
3. Selezionare **Valori percentile** per specificare i valori per calcolare il percentile.
4. Selezionare **Metodo percentile** per selezionare il metodo di calcolo percentile.
5. Facoltativamente, è possibile:
   - Fare clic su **Mancante** per controllare il trattamento dei valori mancanti.
   - Fare clic su **Bootstrap** per ricavare stime affidabili degli errori standard e degli intervalli di confidenza per le stime dei percentili.

Valori mancanti percentili

Valori mancanti
Consente di controllare la modalità di elaborazione dei valori mancanti.

Escludi casi listwise

I casi con valori mancanti per qualsiasi variabile specificata vengono esclusi da tutte le analisi. È l'impostazione predefinita.

Per variabile
Escludi casi con valori mancanti su base variabile per variabile.

Esplora

La procedura Esplora produce statistiche di riepilogo e visualizzazioni grafiche per tutti i casi o per singoli gruppi di casi. Risulta inoltre utile per numerose operazioni, ovvero screening dei dati, identificazione dei valori anomali, descrizione, verifica delle ipotesi e caratterizzazione delle differenze tra sottopopolazioni (gruppi di casi). Lo screening dei dati può evidenziare la presenza di valori insoliti, intervalli vuoti tra i dati o altri elementi specifici. L'esplorazione dei dati può consentire di determinare l'idoneità delle tecniche statistiche selezionate per l'analisi dei dati. L'esplorazione può evidenziare la necessità di eseguire una trasformazione dati se una particolare tecnica richiede una distribuzione normale. In alternativa è possibile utilizzare test non parametrici.

Esempio. Si consideri la distribuzione dei tempi in cui quattro gruppi di ratti imparano a uscire da un labirinto. Per ciascuno dei quattro gruppi, è possibile verificare se la distribuzione dei tempi è approssimativamente normale e se i quattro valori di varianza sono uguali. È inoltre possibile identificare

i casi con i cinque tempi più lunghi e i cinque tempi più brevi. I grafici a scatole e i grafici ramo-foglia riassumono graficamente la distribuzione dei tempi di apprendimento per ciascun gruppo.

Statistiche e grafici. Media, mediana, media ritagliata al 5%, errore standard, varianza, deviazione standard, valore minimo e massimo, intervallo, intervallo interquartile, asimmetria e curtosi e i relativi errori standard, intervallo di confidenza per la media (e il livello di confidenza specificato), percentili, stimatore M di Huber, stimatore M di Andrew, stimatore M decrescente di Hampel, stimatore di Tukey a doppio peso, i cinque valori maggiori e i cinque valori minori, il test di Kolmogorov-Smirnov con il livello di significatività di Lilliefors per il test della normalità e il test di Shapiro-Wilk. Grafici a scatole, grafici ramo-foglia, istogrammi, grafici di normalità e grafici di diffusione contro intensità con test di Levene e trasformazioni.

Considerazioni sui dati relative alla procedura Esplora

Dati. La procedura Esplora può essere utilizzata per le variabili quantitative (livello di misurazione per intervallo o per rapporto). La variabile fattore, utilizzata per suddividere i dati in gruppi di casi, deve includere un numero ragionevole di valori distinti (categorie). Tali valori possono essere stringhe corte o numerici. La variabile etichetta di caso, utilizzata per etichettare i valori anomali in grafici a scatole, può essere una variabile stringa corta, stringa lunga (i primi 15 byte) o numerica.

Ipotesi. La distribuzione dei dati non deve essere necessariamente simmetrica o normale.

Per esplorare i dati

1. Dai menu, scegliere:

   Analizza > **Statistiche descrittive** > **Esplora...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili dipendenti.

Facoltativamente, è possibile:

- Selezionare una o più variabili fattore i cui valori definiranno i gruppi di casi.
- Selezionare una variabile di identificazione per etichettare i casi.
- Fare clic su **Statistiche** per ottenere stimatori robusti, valori anomali, percentili e tabelle delle frequenze.
- Fare clic su **Grafici** per ottenere istogrammi, grafici e test di probabilità normale e grafici di diffusione contro intensità con test di Levene.
- Fare clic su **Opzioni** per ottenere il trattamento dei valori mancanti.

Esplora: Statistica

Descrittive. Queste misure di tendenza centrale e di dispersione vengono visualizzate per impostazione predefinita. Le misure di tendenza centrale indicano l'ubicazione della distribuzione e includono la media, la mediana e la media ritagliata al 5%. Le misure di dispersione mostrano la dissimilarità dei valori e includono errore standard, varianza, deviazione standard, valore minimo e massimo, intervallo e intervallo interquartile. Le statistiche descrittive includono anche le misure della forma della distribuzione; l'asimmetria e la curtosi vengono visualizzate con i rispettivi errori standard. Viene visualizzato anche l'intervallo di confidenza al 95% per la media. È possibile specificare un diverso livello di confidenza.

Stimatori M. Alternative valide alla media e alla mediana del campione per la valutazione dell'ubicazione. Gli stimatori calcolati differiscono nei pesi che si applicano ai casi. Verranno visualizzati lo stimatore M di Huber, lo stimatore M di Andrews, lo stimatore M decrescente di Hampel e lo stimatore di Tukey a doppio peso.

Valori anomali. Consente di visualizzare i cinque valori maggiori e i cinque valori minori con le etichette dei casi.

Percentili

La procedura Percentili visualizza tabelle di percentili. Quando si seleziona **"Percentile"**, vengono visualizzati per impostazione predefinita i valori relativi al 5°, 10°, 25°, 50°, 75°, 90° e 95° percentile. Valori di una variabile quantitativa che suddividono i dati ordinati in due gruppi in modo da visualizzare una percentuale sopra e una sotto.

Quartili

I quartili (il 25°, 50° e 75° percentile) suddividono le osservazioni in quattro gruppi di dimensioni uguali.

Personalizzato

Quando si seleziona **Personalizzato**, immettere almeno un valore per eseguire l'analisi. I valori di input devono essere un numero compreso tra 0 e 100. Utilizzare i pulsanti **Aggiungi**, **Modifica**e **Rimuovi** per gestire i valori nell'elenco dei valori percentili.

Metodo

Per impostazione predefinita, il metodo HAVERAGE è selezionato per calcolare i percentili.

Esplora: Grafici

Grafici a scatole. Queste alternative controllano la visualizzazione dei grafici a scatole quando sono presenti più variabili dipendenti. **Livelli dei fattori insieme** consente di generare una visualizzazione distinta per ciascuna variabile dipendente. All'interno della visualizzazione, vengono visualizzati grafici a scatole per ciascun gruppo definito da una variabile fattore. **Dipendenti insieme** consente di generare una visualizzazione distinta per ciascun gruppo definito da una variabile fattore. All'interno della visualizzazione compaiono grafici a scatole affiancati per ciascuna variabile dipendente. Questo tipo di grafico risulta particolarmente utile quando le singole variabili rappresentano una caratteristica misurata in tempi diversi.

Descrittive. Nel gruppo Descrittive è possibile scegliere grafici ramo-foglia e istogrammi.

Grafici di normalità con test. Consente di visualizzare grafici delle probabilità normale e grafici delle probabilità normale detrendizzati. Viene visualizzato il test di Kolmogorov-Smirnov con un livello di significatività di Lilliefors per il test della normalità. Se sono specificati pesi non interi, la statistica di Shapiro-Wilk viene calcolata quando la dimensione del campione pesato è compresta tra 3 e 50. Per i pesi interi o in assenza di pesi, la statistica viene calcolata quando la dimensione del campione pesato è compresa tra 3 e 5.000.

Confronto tra diffusione e livello con test di Levene. Consente di controllare la trasformazione dati per i grafici di diffusione contro intensità. Per tutti i grafici di diffusione contro intensità vengono visualizzati la pendenza della linea di regressione e i test di Levene per l'omogeneità della varianza. Se si seleziona una trasformazione, i test di Levene si baseranno sui dati trasformati. Se non viene selezionata una variabile fattore, non verranno creati grafici di diffusione contro intensità. **Stima potenza** traccia i logaritmi naturali degli intervalli interquartili verso i logaritmi naturali delle mediane di tutte le celle e inoltre una stima della potenza necessaria per trasformare i dati in modo da raggiungere varianze uguali in tutte le celle. Un grafico diffusione contro intensità consente di identificare la potenza di una trasformazione per stabilizzare (rendere maggiormente uguale) le varianze nei vari gruppi. **Trasformata** consente di selezionare un valore di potenza alternativo, seguendo o meno le indicazioni della stima di potenza, e di produrre i grafici dei dati trasformati. L'intervallo interquartile e la media dei dati trasformati verranno tracciati in un grafico. **Invarianza** consente di ottenere grafici relativi ai dati grezzi. Equivale a una trasformazione con potenza 1.

Esplora: potenza necessaria per la trasformazione dati

Si tratta delle trasformazioni di potenza per i grafici di diffusione contro intensità. Per trasformare i dati è necessario selezionare la potenza corrispondente. È possibile scegliere una delle seguenti opzioni:

- **Logaritmo naturale.** Trasformazione di un logaritmo naturale. È l'impostazione predefinita.
- **1/radice quadrata.** Per ciascun valore dei dati viene calcolato il reciproco della radice quadrata.
- **Reciproco.** Viene calcolato il reciproco di ciascun valore dei dati.
- **Radice quadrata.** Viene calcolata la radice quadrata di ciascun valore dei dati.

- **Quadrato.** Ciascun valore dei dati viene elevato al quadrato.
- **Cubo.** Ciascun valore dei dati viene elevato al cubo.

Esplora: Opzioni

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Escludi casi listwise.** I casi con valori mancanti per qualsiasi variabile dipendente o fattore verranno esclusi da tutte le analisi. È l'impostazione predefinita.
- **Escludi casi pairwise.** I casi che non contengono valori mancanti per le variabili di un gruppo (cella) verranno inclusi nell'analisi per tale gruppo. Il caso può includere valori mancanti per le variabili utilizzate in altri gruppi.
- **Report valori.** I valori mancanti per le variabili fattore vengono trattati come categoria distinta. Tutto l'output viene prodotto per questa categoria supplementare. Le tabelle delle frequenze includono categorie per i valori mancanti. I valori mancanti per una variabile fattore vengono inclusi, ma etichettati come mancanti.

Funzioni aggiuntive del comando EXAMINE

La procedura Esplora usa la sintassi del comando EXAMINE. Il linguaggio della sintassi dei comandi consente inoltre di:

- Richiedere l'output totale e i grafici oltre all'output e ai grafici per i gruppi definiti dalle variabili fattore (con il sottocomando TOTAL).
- Specificare una scala comune per un gruppo di grafici a scatole (con il sottocomando SCALE).
- Specificare le interazioni delle variabili fattore (con il sottocomando VARIABLES).
- Specificare percentili diversi da quelli predefiniti (con il sottocomando PERCENTILES).
- Calcolare i percentili utilizzando uno dei cinque metodi (con il sottocomando PERCENTILES).
- Specificare una trasformazione di potenza per i grafici di diffusione vs. intensità (con il sottocomando PLOT).
- Specificare il numero di valori estremi da visualizzare (con il sottocomando STATISTICS).
- Specificare i parametri per gli stimatori M e gli stimatori robusti di ubicazione (con il sottocomando MESTIMATORS).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Tabelle di contingenza

La procedura Tabelle di contingenza consente di formare tabelle a due vie e a più dimensioni e fornisce una serie di test e misure di associazione per le tabelle a due vie. Il test o la misura da utilizzare vengono determinati in base alla struttura della tabella e al fatto che le categorie siano ordinate o meno.

Ad eccezione dei coefficienti di gamma parziale, le statistiche di Crosstabs e le misure di associazione sono calcolate separatamente per ogni tavolo a due vie. Se si specifica una riga, una colonna o uno strato (variabile di controllo), verrà visualizzato un pannello contenente le statistiche associate e le misurazioni per ciascun valore dello strato (o una combinazione di valori per due o più variabili di controllo). Ad esempio, se la variabile *sesso* è uno strato per la tabella della variabile *coniugato* (sì, no) rispetto alla variabile *tipo di vita* (ottima, soddisfacente, non soddisfacente), i risultati per la tabella a due vie per le donne vengono elaborati separatamente da quelli per gli uomini e quindi stampati come pannelli in successione.

Esempio. È possibile che i clienti rappresentati da piccole società siano più remunerativi per la vendita di servizi (per esempio addestramenti e consulenze) rispetto ai clienti rappresentati da società di grandi dimensioni? Mediante una tavola di contingenza è possibile scoprire che la maggior parte delle società di piccole dimensioni (con un numero di dipendenti inferiore a 500) fruttano alti profitti per i servizi, mentre la maggior parte delle grandi società (con oltre 2,500 dipendenti) fruttano profitti di scarsa entità.

Statistiche e misure di associazione. chi-quadrato di Pearson, chi-quadrato del rapporto di verosimiglianza, test di associazione lineare per lineare, test esatto di Fisher, chi-quadrato corretto di Yates, *r* di Pearson, rho di Spearman, coefficiente di contingenza, phi, *V* di Cramér, lambda simmetrico e asimmetrico, tau di Goodman e Kruskal, coefficiente di incertezza, gamma, *d* di Somers, tau-*b* di Kendall, tau-*c* di Kendall, coefficiente eta, kappa di Cohen, stima del rischio relativo, rapporto odd, test di McNemar, statistiche di Cochran e Mantel-Haenszel e statistiche delle proporzioni di colonna.

Considerazioni sui dati relativi alle tabelle di contingenza

Dati. Per definire le categorie di ciascuna variabile di tabella, utilizzare i valori di una variabile numerica o stringa. Ad esempio, per la variabile *sesso*, è possibile codificare i dati come 1 e 2 oppure come *maschio* e *femmina*.

Ipotesi. Alcune statistiche e misure assumono categorie ordinate (dati ordinali) o valori quantitativi (dati misurati per intervallo o per rapporto), come indicato nella sezione sulle statistiche. Se le variabili della tabella prevedono categorie non ordinate (dati nominali), sono disponibili altri valori validi. Per le statistiche basate sul chi-quadrato (phi, *V* di Cramér e coefficiente di contingenza), i dati devono essere un campione casuale di una distribuzione multinomiale.

Nota: le variabili ordinali possono essere codici numerici che rappresentano le categorie (ad esempio, 1 = *basso*, 2 = *medio*, 3 = *alto*) o valori stringa. Si suppone tuttavia che l'ordine alfabetico dei valori di stringa rifletta l'esatto ordine delle categorie. Ad esempio, per una variabile stringa con i valori *basso*, *medio*, *alto*, l'ordine delle categorie viene interpretato come *alto*, *basso*, *medio*, ma questo non è l'ordine corretto. In generale, per rappresentare i dati ordinali, è più sicuro utilizzare i codici numerici.

Per ottenere tavole di contingenza

1. Dai menu, scegliere:

   Analizza > **Statistiche descrittive** > **Tabelle di contingenza...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili di riga e una o più variabili di colonna.

Facoltativamente, è possibile:

- Selezionare una o più variabili di controllo.
- Fare clic su **Statistiche** per ottenere test e misure di associazione per tabelle o sottotabelle a due vie.
- Fare clic su**Celle** per ottenere valori, percentuali e residui osservati e attesi.
- Fare clic su**Formato** per controllare l'ordine delle categorie.

Strati delle tabelle di contingenza

Se vengono selezionate una o più variabili di strato, verrà prodotta una tavola di contingenza distinta per ciascuna categoria di ciascuna variabile di strato (variabile di controllo). Ad esempio, se si dispone di una variabile di riga, una variabile di colonna e una variabile di strato con due categorie, si otterrà una tabella a due vie per ciascuna categoria della variabile di strato. Per creare un altro strato di variabili di controllo, fare clic su **Successivo**. Verranno create sottotabelle per ogni combinazione delle categorie di ciascuna variabile del primo strato con ciascuna variabile del secondo e così via. Se sono richieste statistiche e misure di associazione, verranno applicate solo alle sottotabelle a due vie.

Grafici a barre raggruppate di tabelle di contingenza

Visualizza grafici a barre raggruppate. Nei grafici a barre raggruppate è possibile riepilogare i dati relativi a gruppi di casi. È disponibile un gruppo di barre per ciascun valore della variabile specificata in Righe. La variabile che definisce le barre contenute in ogni gruppo è quella specificata in Colonne. Per ciascun valore della variabile è disponibile un insieme di barre con colori e modelli diversi. Se in Colonne o Righe si specificano più variabili, verrà prodotto un grafico a barre raggruppate per ciascuna combinazione delle due variabili.

Tabelle di contingenza che visualizzano le variabili di strato negli strati della tabella

Visualizza variabili di strato negli strati della tabella. È possibile scegliere di visualizzare le variabili di strato (variabili di controllo) come strati della tabella nella tabella delle tavole di contingenza. Ciò consente di creare delle viste che mostrano le statistiche generali per le variabili di riga e colonna e che consentono di eseguire il drill-down delle categorie delle variabili di strato.

Un esempio che utilizza il file di dati *demo.sav* (disponibile nella cartella dei campioni della cartella di installazione) viene visualizzato qui di seguito ed è stato ottenuto nel seguente modo:

1. Selezionare *Categoria di reddito in migliaia (inccat)* come variabile della riga, *Possiede PDA (ownpda)* come variabile della colonna e *Livello di istruzione (ed)* come variabile dello strato.
2. Selezionare **Visualizza variabili di strato negli strati della tabella**.
3. Selezionare **Colonna** nella finestra di dialogo secondaria Visualizzazione cella.
4. Eseguire la procedura Tabelle di contingenza, fare doppio clic sulla tabella delle tavole di contingenza e selezionare **Diploma di laurea** dall'elenco a discesa Livello di istruzione.

La vista selezionata della tabella delle tavole di contingenza mostra le statistiche per i rispondenti che hanno un diploma di laurea.

Statistiche delle tabelle di contingenza

Chi-quadrato. Per tabelle con due righe e due colonne, scegliere **Chi-quadrato** per calcolare il chi-quadrato di Pearson, il chi-quadrato del rapporto di verosimiglianza, il test esatto di Fisher e il chi-quadrato corretto di Yates (correzione di continuità). Per le tabelle 2 × 2, il test esatto di Fisher viene calcolato quando una tabella che non deriva da righe o colonne mancanti in una tabella più grande ha una cella con una frequenza prevista inferiore a 5. Yates ' corretto chi - quadrato viene calcolato per tutte le altre 2 × 2 tavole. Per tabelle con un numero qualsiasi di righe e colonne, selezionare **Chi-quadrato** per calcolare il chi-quadrato di Pearson e il chi-quadrato del rapporto di verosimiglianza. Se entrambe le variabili delle tabelle sono quantitative, l'opzione **Chi-quadrato** restituisce il test dell'associazione lineare.

Correlazioni. Per le tabelle in cui sia le righe che le colonne contengono valori ordinati, l'opzione **Correlazioni** restituisce il coefficiente di correlazione di Spearman, rho (solo dati numerici). Il coefficiente rho di Spearman è una misura di associazione tra punteggi di rango. Se entrambe le variabili delle tabelle (fattori) sono quantitative, **Correlazioni** restituisce il coefficiente di correlazione di Pearson, *r*, una misura dell'associazione lineare tra le variabili.

Nominale. Per i dati nominali (nessun ordine intrinseco, ad esempio cattolico, protestante o ebreo), è possibile selezionare **Coefficiente di contingenza**, **Phi** (coefficiente) **e V di Cramér**, **Lambda** (lambda simmetrici e asimmetrici e tau di Goodman e Kruskal) e **Coefficiente di incertezza**.

- *Coefficiente di contingenza.*.Una misura di associazione basata sul chi-quadrato. Questo coefficiente è sempre compreso tra 0 e 1, dove 0 indica nessuna associazione tra le variabili di riga e colonna e i valori vicini a 1 indicano un alto grado di associazione tra le variabili. Il valore massimo possibile dipende dal numero di righe e colonne in una tabella.
- *Phi e V di Cramer*. Phi è una misura di associazione basata sul chi - quadrato che implica la divisione della statistica chi - quadrato per la dimensione del campione e la radice quadrata del risultato. V di Cramer è una misura di associazione basata sul chi-quadrato.
- *Lambda*. Una misura di associazione che riflette la riduzione proporzionale nell'errore quando i valori della variabile indipendente vengono utilizzati per prevedere i valori della variabile dipendente. Un valore pari a 1 significa che la variabile indipendente stima perfettamente la variabile dipendente. Un valore pari a 0 significa che la variabile indipendente non è di alcun aiuto nella stima della variabile dipendente.
- *Coefficiente di incertezza.*.Una misura di associazione che indica la riduzione proporzionale nell'errore quando i valori di una variabile vengono utilizzati per prevedere i valori dell'altra variabile. Un valore di 0,83, ad esempio, indica che la conoscenza di una variabile riduce dell'83% l'errore nella stima dei

valori dell'altra variabile. Il programma calcola sia la versione simmetrica che quella asimmetrica del coefficiente di incertezza.

Ordinale. Per tabelle in cui sia le righe che le colonne contengono valori ordinati, selezionare **Gamma** (gamma di ordine zero per tabelle a 2 vie e gamma condizionali per tabelle da 3 a 10 vie), **Tau-b di Kendall** e **Tau-c di Kendall**. Per desumere le categorie delle colonne delle righe, selezionare **D di Somers**.

- *Gamma*. Una misura simmetrica dell'associazione tra due variabili ordinali che varia tra -1 e 1. Valori vicini a un valore assoluto di 1 indicano una forte relazione tra le due variabili. Valori prossimi allo zero indicano scarsità o assenza di relazione. In caso di tabelle a 2 vie verranno visualizzati gamma di ordine zero. Per da 3 vie a n vie, vengono visualizzati i gamma condizionali.
- *D di Somers*. Misura di associazione tra due variabili ordinali che va da -1 a 1. Valori vicini a un valore assoluto di 1 indicano una forte relazione tra le due variabili e i valori vicini a 0 indicano poco o nessun rapporto tra le variabili. È una estensione asimmetrica di gamma dalla quale differisce solo per l'inclusione del numero di coppie non correlate nella variabile indipendente. Viene calcolata anche una versione simmetrica di questa statistica.
- *Tau-b di Kendall*. Una misura non parametrica di correlazione per le variabili ordinali o classificate che tiene conto delle correlazioni. Il segno del coefficiente indica la direzione della correlazione e il valore assoluto la sua forza. Valori assoluti maggiori indicano correlazioni maggiori. I valori possibili variano da -1 a +1, ma il valore -1 o +1 può solo essere ottenuto da tabelle quadrate.
- *Tau-c di Kendall*. Una misura non parametrica di associazione per variabili ordinali che ignora le correlazioni. Il segno del coefficiente indica la direzione della correlazione e il valore assoluto la sua forza. Valori assoluti maggiori indicano correlazioni maggiori. I valori possibili variano da -1 a +1, ma il valore -1 o +1 può solo essere ottenuto da tabelle quadrate.

Nominale per intervallo. Se una variabile è categoriale e l'altra quantitativa, scegliere **Eta**. La variabile categoriale deve essere codificata numericamente.

- *Eta*. Una misura di associazione che varia da 0 a 1, con 0 che indica nessuna associazione tra le variabili riga e colonna e valori prossimi a 1 che indicano un alto livello di associazione. Eta è appropriata per una variabile dipendente misurata su una scala per intervallo e una variabile indipendente con numero limitato di categorie. Vengono calcolati due valori eta: il primo tratta la variabile riga come variabile di intervallo e l'altro tratta la variabile colonna come variabile di intervallo.

Kappa. Il kappa di Cohen misura l'accordo tra le valutazioni di due valutatori quando entrambi valutano lo stesso oggetto. Un valore pari a 1 indica accordo perfetto. Un valore pari a 0 indica che l'accordo può essere considerato casuale. Il kappa è basato su una tabella quadrata in cui i valori di riga e di colonna rappresentano la stessa scala. Qualsiasi cellula che abbia osservato valori per una variabile ma non l'altra è assegnata un conteggio di 0. Kappa non viene calcolato se il tipo di archiviazione dei dati (stringa o numerica) non è uguale per le due variabili. Nel caso della variabile stringa, entrambe le variabili devono avere la stessa lunghezza definita.

Rischio.Per le tabelle 2 x 2, una misura dell'intensità dell'associazione tra la presenza di un fattore e la ricorrenza di un evento. Un valore pari a 1 indica che il fattore non è associato all'evento. Il rapporto odd può essere usato come una stima del rischio relativo quando la ricorrenza del fattore è rara.

McNemar. Un test non parametrico per due variabili dicotomiche correlate. Verifica la presenza di cambiamenti nelle risposte mediante la distribuzione chi-quadrato. Il test è molto utile in disegni sperimentali del tipo 'prima e dopo', per rilevare cambiamenti di risposta. Per tabelle quadrate di dimensioni maggiori, viene notificato il test della simmetria di McNemar-Bowker.

Statistiche di Cochran e Mantel-Haenszel. Le statistiche di Cochran e Mantel - Haenszel possono essere utilizzate per verificare l'indipendenza tra una variabile fattore dicotomica e una variabile di risposta dicotomica, condizionali ai modelli di covariata definiti da una o più variabili di livello (controllo). Si noti che mentre le altre statistiche vengono calcolate strato per strato, le statistiche di Cochran e Mantel-Haenszel vengono calcolate una volta per tutti gli strati.

Visualizzazione celle delle tabelle di contingenza

Per facilitare l'individuazione di modelli di dati che danno origine a un test del chi-quadrato significativo, la procedura per le tabelle di contingenza visualizza le frequenze attese e tre tipi di residui (devianze) che

misurano la differenza tra le frequenze osservate e quelle attese. Ogni cella della tabella può contenere qualsiasi combinazione dei conteggi, delle percentuali e dei residui selezionati.

Conteggi. Il numero di casi effettivamente osservati e il numero di casi attesi se le variabili di riga e di colonna sono reciprocamente indipendenti. È possibile scegliere di nascondere i conteggi che sono inferiori a un numero intero specificato. I valori nascosti verranno visualizzati come **< N**, dove **N** è il numero intero specificato. Il numero intero specificato deve essere superiore o uguale a 2, ma è consentito utilizzare il valore 0 per specificare che non viene nascosto alcun conteggio.

Confronta proporzioni di colonna. Questa opzione calcola i confronti pairwise delle proporzioni di colonna e indica quali coppie di colonne (per una determinata riga) sono differenti in modo significativo. Le differenze significative vengono indicate nella tabella delle tavole di contingenza con formattazione in stile APA utilizzando lettere come pedici e vengono calcolate al livello di significatività 0.05. *Nota*: se questa opzione viene specificata senza selezionare i conteggi osservati o le percentuali di colonna, i conteggi osservati vengono inclusi nella tabella delle tavole di contingenza con lettere come pedice in stile APA indicanti i risultati dei test delle proporzioni di colonna.

- **Adatta i valori P (metodo di Bonferroni).** I confronti pairwise delle proporzioni di colonna utilizzano la correzione di Bonferroni, che adatta il livello di significatività osservato per il fatto che vengono eseguiti confronti multipli.

Percentuali. Le percentuali possono essere aggiunte nelle righe o nelle colonne. Sono disponibili anche le percentuali del numero totale di casi rappresentati nella tabella (un solo strato).

Nota: Se **Nascondi piccoli conteggi** viene selezionato nel gruppo Contenuti, anche le percentuali associate a conteggi nascosti sono nascoste.

Residui. I residui grezzi non standardizzati forniscono la differenza tra valori osservati e attesi. Sono inoltre disponibili residui standardizzati e standardizzati adattati.

- *Non standardizzati*.La differenza tra un valore osservato e il valore previsto. Per valore atteso si intendo il numero di casi atteso nella cella in assenza di relazione tra le due variabili. Un residuo positivo indica che ci sono più casi nella cella di quanti ce ne sarebbero se le variabili di riga e di colonna fossero indipendenti.
- *Standardizzati*.Il residuo diviso per una stima della sua deviazione standard. I residui standardizzati, noti anche come residui di Pearson, hanno una media pari a 0 e una deviazione standard pari a 1.
- *Standardizzati adattati*.Il residuo per una cella (valore osservato meno valore previsto) diviso per una stima del proprio errore standard. Il residuo standardizzato risultante è espresso in unità di deviazione standard sopra o sotto la media.

Crea tabella di stile APA. Crea tabelle di output che aderiscono alle linee guida in stile APA.

Nota: Le Opzioni **Osservate**, **Attese**, **Row**, **Colonna**e **Totale** non sono disponibili quando **Crea APA la tabella di stile** è selezionata.

Pesi non interi. I conteggi delle celle in genere sono valori interi, in quanto rappresentano il numero di casi in ogni cella. Se, tuttavia, il file di dati è attualmente pesato in base a una variabile peso con valori frazionari (ad esempio, 1,25), i conteggi delle celle possono essere espressi anche in valori frazionari. È possibile troncare o arrotondare i valori prima o dopo aver calcolato i conteggi delle celle oppure utilizzare conteggi delle celle frazionari per la visualizzazione delle tabelle e dei calcoli statistici.

- *Arrotonda frequenze cella*. I pesi dei casi vengono usati come tali, mentre i pesi accumulati nelle celle vengono arrotondati prima di calcolare qualsiasi statistica.
- *Tronca conteggi cella*. I pesi dei casi vengono usati come tali, mentre i pesi accumulati nelle celle vengono troncati prima di calcolare qualsiasi statistica.
- *Arrotonda pesi caso*. I pesi dei casi vengono arrotondati prima dell'uso.
- *Tronca pesi del caso*. I pesi dei casi sono troncati prima dell'uso.
- *Nessun adattamento*.I pesi dei casi vengono utilizzati così come sono e vengono utilizzati i conteggi delle celle frazionali. Tuttavia, quando sono necessarie statistiche esatte (disponibili solo con Campionamento e Test), i pesi accumulati nelle celle vengono troncati o arrotondati prima di calcolare le statistiche di test esatte.

Formato della tabella di tavole di contingenza

È possibile disporre le righe nell'ordine crescente o decrescente dei valori della variabile di riga.

Riepiloga

La procedura Riassumi consente di calcolare la statistica del sottogruppo per le variabili all'interno delle categorie di una o più variabili di raggruppamento. Tutti i livelli della variabile di raggruppamento vengono incrociati. È possibile scegliere l'ordine in cui vengono visualizzate le statistiche. Per ciascuna variabile di tutte le categorie verranno inoltre visualizzate le statistiche di riepilogo. I valori dei dati di ciascuna categoria possono essere inseriti nell'elenco o eliminati, ma nei dataset di grandi dimensioni, è possibile scegliere di elencare solo i primi *n* casi.

Esempio. Qual è l'importo medio delle vendite per area e industria del cliente? Si potrebbe scoprire che l'importo medio delle vendite è leggermente superiore nell'area occidentale rispetto alle altre aree e che ai clienti di quest'area è associato l'importo medio più alto.

Statistiche. Somma, numero di casi, media, mediana, mediana raggruppata, errore standard della media, minimo, massimo, intervallo, valore della variabile della prima categoria della variabile di raggruppamento, valore della variabile dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del *numero* totale, percentuale della somma in, percentuale del *numero di casi* in, media geometrica e media armonica.

Considerazioni sui dati relative alla procedura Riassumi

Dati. Le variabili di raggruppamento sono variabili categoriali che possono contenere valori stringa o numerici. Il numero di categorie dovrebbe essere limitato. Le altre variabili dovrebbero essere classificabili.

Ipotesi. Alcune delle statistiche del sottogruppo facoltative, quali la media e la deviazione standard, sono basate sulla teoria della normalità e sono idonee per le variabili quantitative con distribuzione simmetrica dei dati. La mediana e l'intervallo sono statistiche robuste, idonee per le variabili quantitative che possono o meno soddisfare l'ipotesi di normalità.

Per ottenere riepiloghi dei casi

1. Dai menu, scegliere:

   Analizza > **Report** > **Riepiloghi dei casi...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili.

Facoltativamente, è possibile:

- Selezionare una o più variabili di raggruppamento per suddividere i dati in sottogruppi.
- Fare clic su **Opzioni** per modificare il titolo dell'output, aggiungere una didascalia al di sotto dell'output o escludere casi con valori mancanti.
- Fare clic su **Statistiche** per visualizzare statistiche facoltative.
- Selezionare **Visualizza casi** per visualizzare un elenco dei casi inclusi in ciascun sottogruppo. Per impostazione predefinita, vengono elencati solo i primi 100 casi nel file. È possibile aumentare o ridurre il valore di **Limita i casi ai primi** *n* oppure deselezionare l'opzione per visualizzare l'elenco di tutti i casi.

Riassumi: Opzioni

SPSS consente di modificare il titolo dell'output o di aggiungere una didascalia che verrà visualizzata sotto alla tabella di output. È possibile controllare gli a capo automatici nei titoli e nelle didascalie digitando \n dove si desidera inserire un'interruzione della riga nel testo.

È inoltre possibile scegliere di visualizzare o eliminare i sottototali e di includere o escludere casi con valori mancanti per qualsiasi variabile utilizzata nelle analisi. È spesso consigliabile contrassegnare nell'output i casi mancanti utilizzando un punto o un asterisco. Immettere un carattere, una frase o codice che si desidera venga visualizzato per indicare che un valore è mancante. In caso contrario, ai casi mancanti non verrà applicato alcun identificativo nell'output.

Riassumi: Statistiche

È possibile scegliere una o più delle seguenti statistiche del sottogruppo per le variabili in ciascuna categoria di ogni variabile di raggruppamento: somma, numero di casi, media, mediana, mediana raggruppata, errore standard della media, minimo, massimo, intervallo, valore della variabile della prima categoria della variabile di raggruppamento, valore della variabile dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del *numero* totale, percentuale della somma in, percentuale del *numero di casi* in, media geometrica, media armonica. L'ordine in cui compaiono le statistiche nell'elenco Statistiche delle celle corrisponde all'ordine in cui verranno visualizzate nell'output. Per ciascuna variabile vengono visualizzate anche le statistiche di riepilogo in tutte le categorie.

Primo. Visualizza il primo valore dei dati incontrato nel file di dati.

Media geometrica. La radice ennesima del prodotto dei valori dei dati, dove n è il numero di casi.

Mediana raggruppata. La mediana calcolata per i dati codificati in gruppi. Ad esempio, con i dati di età, se ogni valore nella 30ina e codificato 35, ogni valore nella 40ina è codificato 45 e così via, la mediana raggruppata è la mediana calcolata dai dati codificati.

media armonica. Utilizzato per stimare una dimensione media del gruppo quando le dimensioni del campione nei gruppi non sono uguali. La media armonica è il numero totale di campioni diviso per la somma dei reciproci delle dimensioni del campione.

Curtosi. Una misura della misura in cui esistono valori anomali. Per la distribuzione normale, il valore della statistica di curtosi è zero. Una curtosi positiva indica che i dati mostrano valori anomali più estremi di una distribuzione normale. La kurtosis negativa indica che i dati esibiano valori anomali meno estremi di una distribuzione normale. La definizione di kurtosi che si usa, dove il valore è di 0 per una distribuzione normale, è a volte indicata come kurtosi in eccesso. Alcuni software possono segnalare kurtosi tale che il valore è di 3 per una distribuzione normale.

Ultimo. Visualizza l'ultimo valore di dati incontrato nel file di dati.

maximum. Il valore più alto di una variabile numerica.

Media. Una misura della tendenza centrale. La media aritmetica, ossia la somma divisa per il numero di casi.

Median. Il valore al di sopra e al di sotto del quale ricade la metà dei casi: il 50° percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori estremamente bassi o alti.

Minimo. Il valore più basso di una variabile numerica.

s . Il numero di casi (osservazioni o record).

Percentuale di N totale. Percentuale del numero totale di casi in ciascuna categoria.

Percentuale della somma totale..Percentuale della somma totale in ciascuna categoria.

Intervallo. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.

Asimmetria. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con asimmetria positiva ha una coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica uno scostamento dalla simmetria.

deviazione standard. Misura della dispersione intorno alla media. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95% dei casi rientra in due deviazioni standard. Se, ad esempio, l'età media è 45 e la deviazione standard è 10, il 95% dei casi sarà compreso tra 25 e 65 in un distribuzione normale.

Errore standard della curtosi. Il rapporto tra la curtosi e il suo errore standard può essere utilizzato come test di normalità (cioè, si può rifiutare la normalità se il rapporto è inferiore a -2 o superiore a +2). Un valore positivo elevato per la curtosi indica che le code della distribuzione sono più lunghe di quelle di una distribuzione normale; un valore negativo per la curtosi indica code più corte (che diventano simili a quelle di una distribuzione uniforme a forma di scatola).

Errore standard della media. Una misura di quanto il valore della media può variare da campione a campione preso dalla stessa distribuzione. Può essere usata per comparare genericamente la media osservata rispetto a un valore ipotizzato (ovvero, è possibile concludere che i due valori sono diversi se il rapporto della differenza rispetto all'errore standard è inferiore a -2 o maggiore di +2).

Errore standard dell'assimetria. Il rapporto tra la skewness e il suo errore standard può essere utilizzato come test di normalità (cioè, si può rifiutare la normalità se il rapporto è inferiore a -2 o superiore a +2). Un valore positivo elevato per l'asimmetria indicata una coda a destra lunga; un valore negativo estremo indica una coda a sinistra lunga.

Somma. La somma o il totale di tutti i valori non mancanti di tutti i casi.

Varianza. Una misura di dispersione intorno alla media, uguale alla somma dei quadrati delle deviazioni dalla media divisa per un valore inferiore al numero di casi. La varianza è espressa in quadrati dell'unità di misura della variabile.

Medie

La procedura Medie consente di calcolare le medie dei sottogruppi e la statistica univariata correlata per le variabili dipendenti all'interno delle categorie di una o più variabili indipendenti. È inoltre possibile ottenere l'analisi della varianza a una via, eta e test di linearità.

Esempio. Misurare la quantità media di grasso assorbita da tre diversi tipi di olii alimentari ed eseguire un'analisi della varianza a una via per verificare se le medie differiscono.

Statistiche. Somma, numero di casi, media, mediana, mediana raggruppata, errore standard della media, minimo, massimo, intervallo, valore della variabile della prima categoria della variabile di raggruppamento, valore della variabile dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del *numero* totale, percentuale della somma in, percentuale del *numero di casi* in, media geometrica e media armonica. Le opzioni includono analisi della varianza, eta, eta al quadrato, e test di linearità *R* e R^2.

Considerazioni sui dati relativi alle medie

Dati. Le variabili dipendenti sono quantitative e le variabili indipendenti sono categoriali. I valori delle variabili categoriali possono essere di tipo numerico o stringa.

Ipotesi. Alcune delle statistiche del sottogruppo facoltative, quali la media e la deviazione standard, sono basate sulla teoria della normalità e sono idonee per le variabili quantitative con distribuzione simmetrica dei dati. La mediana è una statistica robusta, idonea per le variabili quantitative che possono o meno soddisfare l'ipotesi di normalità. L'analisi della varianza è robusta per quanto riguarda le alterazioni della normalità, ma i dati in ciascuna cella devono essere simmetrici. L'analisi della varianza assume inoltre che i gruppi provengano da popolazioni con valori di varianza uguali. Per verificare questa ipotesi, utilizzare il test di omogeneità della varianza di Levene, disponibile nella procedura ANOVA a una via.

Per ottenere le medie dei sottogruppi

1. Dai menu, scegliere:

   Analizza > **Confronta medie** > **Medie...**

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una o più variabili dipendenti.
3. Usare uno dei seguenti metodi per selezionare le variabili categoriali indipendenti:
   - Selezionare una o più variabili indipendenti. Per ciascuna variabile indipendente vengono visualizzati risultati distinti.
   - Selezionare uno o più strati di variabili indipendenti. Ogni strato suddivide ulteriormente il campione. Se è presente una sola variabile indipendente nello Strato 1 e una sola nello Strato 2, i risultati verranno visualizzati in una tabella incrociata e non in tabelle distinte per ciascuna variabile indipendente.
4. Oppure fare clic su **Opzioni** per ottenere statistiche facoltative, analisi della tabella di varianza, età, età quadrato, *R* e R^2.

Medie: Opzioni

È possibile scegliere una o più delle seguenti statistiche del sottogruppo per le variabili in ciascuna categoria di ogni variabile di raggruppamento: somma, numero di casi, media, mediana, mediana raggruppata, errore standard della media, minimo, massimo, intervallo, valore della variabile della prima categoria della variabile di raggruppamento, valore della variabile dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del *numero* totale, percentuale della somma in, percentuale del *numero di casi* in, media geometrica e media armonica. È possibile modificare l'ordine in cui compare la statistica del sottogruppo. L'ordine in cui compaiono le statistiche nell'elenco Statistiche delle celle corrisponde all'ordine in cui verranno visualizzate nell'output. Per ciascuna variabile vengono visualizzate anche le statistiche di riepilogo in tutte le categorie.

Primo. Visualizza il primo valore dei dati incontrato nel file di dati.

Media geometrica. La radice ennesima del prodotto dei valori dei dati, dove n è il numero di casi.

Mediana raggruppata. La mediana calcolata per i dati codificati in gruppi. Ad esempio, con i dati di età, se ogni valore nella 30ina e codificato 35, ogni valore nella 40ina è codificato 45 e così via, la mediana raggruppata è la mediana calcolata dai dati codificati.

media armonica. Utilizzato per stimare una dimensione media del gruppo quando le dimensioni del campione nei gruppi non sono uguali. La media armonica è il numero totale di campioni diviso per la somma dei reciproci delle dimensioni del campione.

Curtosi. Una misura della misura in cui esistono valori anomali. Per la distribuzione normale, il valore della statistica di curtosi è zero. Una curtosi positiva indica che i dati mostrano valori anomali più estremi di una distribuzione normale. La kurtosis negativa indica che i dati esibiano valori anomali meno estremi di una distribuzione normale. La definizione di kurtosi che si usa, dove il valore è di 0 per una distribuzione normale, è a volte indicata come kurtosi in eccesso. Alcuni software possono segnalare kurtosi tale che il valore è di 3 per una distribuzione normale.

Ultimo. Visualizza l'ultimo valore di dati incontrato nel file di dati.

maximum. Il valore più alto di una variabile numerica.

Media. Una misura della tendenza centrale. La media aritmetica, ossia la somma divisa per il numero di casi.

Median. Il valore al di sopra e al di sotto del quale ricade la metà dei casi: il 50° percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori estremamente bassi o alti.

Minimo. Il valore più basso di una variabile numerica.

s . Il numero di casi (osservazioni o record).

Percentuale del numero di casi totale. Percentuale del numero totale di casi in ciascuna categoria.

Percentuale della somma totale. Percentuale della somma totale in ciascuna categoria.

Intervallo. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.

Asimmetria. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con asimmetria positiva ha una coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica uno scostamento dalla simmetria.

deviazione standard. Misura della dispersione intorno alla media. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95% dei casi rientra in due deviazioni standard. Se, ad esempio, l'età media è 45 e la deviazione standard è 10, il 95% dei casi sarà compreso tra 25 e 65 in un distribuzione normale.

Errore standard della curtosi. Il rapporto tra la curtosi e il suo errore standard può essere utilizzato come test di normalità (cioè, si può rifiutare la normalità se il rapporto è inferiore a -2 o superiore a +2). Un valore positivo elevato per la curtosi indica che le code della distribuzione sono più lunghe di quelle di una distribuzione normale; un valore negativo per la curtosi indica code più corte (che diventano simili a quelle di una distribuzione uniforme a forma di scatola).

Errore standard della media. Una misura di quanto il valore della media può variare da campione a campione preso dalla stessa distribuzione. Può essere usata per comparare genericamente la media osservata rispetto a un valore ipotizzato (ovvero, è possibile concludere che i due valori sono diversi se il rapporto della differenza rispetto all'errore standard è inferiore a -2 o maggiore di +2).

Errore standard dell'assimetria. Il rapporto tra la skewness e il suo errore standard può essere utilizzato come test di normalità (cioè, si può rifiutare la normalità se il rapporto è inferiore a -2 o superiore a +2). Un valore positivo elevato per l'asimmetria indicata una coda a destra lunga; un valore negativo estremo indica una coda a sinistra lunga.

Somma. La somma o il totale di tutti i valori non mancanti di tutti i casi.

Varianza. Una misura di dispersione intorno alla media, uguale alla somma dei quadrati delle deviazioni dalla media divisa per un valore inferiore al numero di casi. La varianza è espressa in quadrati dell'unità di misura della variabile.

Statistiche per il primo strato

Tabella ANOVA ed Eta.Visualizza una tabella di analisi della varianza unidirezionale e calcola le misure di associazione eta ed eta al quadrato per ogni variabile indipendente del primo livello.

Test di linearità.Calcola la somma dei quadrati, i gradi di libertà e la media dei quadrati associati alle componenti lineari e non lineari, nonché il rapporto F, R e R-quadrato. La linearità non viene calcolata se la variabile indipendente è una stringa corta.

Cubi OLAP

La procedura Cubi OLAP (Online Analytical Processing) consente di calcolare i totali, le medie e le altre statistiche univariate per le variabili di riepilogo continue all'interno delle categorie di una o più variabili di raggruppamento categoriali. Nella tabella viene creato uno strato distinto per ciascuna categoria di ogni variabile di raggruppamento.

Esempio. Le vendite totali e medie di diverse aree e le linee di prodotti all'interno delle aree.

Statistiche. Somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del numero di casi totale, percentuale della somma totale entro variabili di raggruppamento, percentuale del numero di casi totale entro variabili di raggruppamento, media geometrica, media armonica.

Considerazioni sui dati dei cubi OLAP

Dati. Le variabili di riepilogo sono quantitative (variabili continue misurate su una scala di intervallo o di rapporto) e le variabili di raggruppamento sono categoriali. I valori delle variabili categoriali possono essere di tipo numerico o stringa.

Ipotesi. Alcune delle statistiche del sottogruppo facoltative, quali la media e la deviazione standard, sono basate sulla teoria della normalità e sono idonee per le variabili quantitative con distribuzione simmetrica dei dati. La mediana e l'intervallo sono statistiche robuste, idonee per le variabili quantitative che possono o meno soddisfare l'ipotesi di normalità.

Per ottenere cubi OLAP

1. Dai menu, scegliere:

   Analizza > **Report** > **Cubi OLAP...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili di riepilogo continue.
3. Selezionare una o più variabili categoriali di raggruppamento

Oppure:

- Selezionare statistiche di riepilogo diverse (fare clic su **Statistiche**). Prima di selezionare le statistiche di riepilogo è necessario selezionare una o più variabili di raggruppamento.
- Calcolare differenze tra coppie di variabili e coppie di gruppi definiti da una variabile di raggruppamento (fare clic su **Differenze**).
- Creare titoli di tabella personalizzati (fare clic su **Titolo**).
- Nascondere i conteggi che sono inferiori a un numero intero specificato. I valori nascosti verranno visualizzati come **< N**, dove **N** è il numero intero specificato. Il numero intero specificato deve essere superiore o uguale a 2.

Cubi OLAP: Statistiche

È possibile scegliere una o più delle seguenti statistiche del sottogruppo per le variabili di riepilogo in ciascuna categoria di ogni variabile di raggruppamento: somma, numero di casi, media, mediana, mediana raggruppata, errore standard della media, minimo, massimo, intervallo, valore della variabile della prima categoria della variabile di raggruppamento, valore della variabile dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale di casi totali, percentuale della somma totale, percentuale dei casi totali entro le variabili di raggruppamento, percentuale della somma totale entro le variabili di raggruppamento, media geometrica e media armonica.

È possibile modificare l'ordine in cui compare la statistica del sottogruppo. L'ordine in cui compaiono le statistiche nell'elenco Statistiche delle celle corrisponde all'ordine in cui verranno visualizzate nell'output. Per ciascuna variabile vengono visualizzate anche le statistiche di riepilogo in tutte le categorie.

Primo. Visualizza il primo valore dei dati incontrato nel file di dati.

Media geometrica. La radice ennesima del prodotto dei valori dei dati, dove n è il numero di casi.

Mediana raggruppata. La mediana calcolata per i dati codificati in gruppi. Ad esempio, con i dati di età, se ogni valore nella 30ina e codificato 35, ogni valore nella 40ina è codificato 45 e così via, la mediana raggruppata è la mediana calcolata dai dati codificati.

media armonica. Utilizzato per stimare una dimensione media del gruppo quando le dimensioni del campione nei gruppi non sono uguali. La media armonica è il numero totale di campioni diviso per la somma dei reciproci delle dimensioni del campione.

Curtosi. Una misura della misura in cui esistono valori anomali. Per la distribuzione normale, il valore della statistica di curtosi è zero. Una curtosi positiva indica che i dati mostrano valori anomali più estremi di una distribuzione normale. La kurtosis negativa indica che i dati esibiano valori anomali meno estremi di una distribuzione normale. La definizione di kurtosi che si usa, dove il valore è di 0 per una distribuzione

normale, è a volte indicata come kurtosi in eccesso. Alcuni software possono segnalare kurtosi tale che il valore è di 3 per una distribuzione normale.

Ultimo. Visualizza l'ultimo valore di dati incontrato nel file di dati.

maximum. Il valore più alto di una variabile numerica.

Media. Una misura della tendenza centrale. La media aritmetica, ossia la somma divisa per il numero di casi.

Median. Il valore al di sopra e al di sotto del quale ricade la metà dei casi: il 50° percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori estremamente bassi o alti.

Minimo. Il valore più basso di una variabile numerica.

s . Il numero di casi (osservazioni o record).

Percentuale del numero di casi in. Percentuale del numero di casi per la variabile di raggruppamento specificata all'interno delle categorie di altre variabili di raggruppamento. Se si ha una sola variabile di raggruppamento, questo valore è identico alla percentuale del numero totale di casi.

Percentuale della somma in.Percentuale della somma per la variabile di raggruppamento specificata all'interno delle categorie delle altre variabili di raggruppamento. Se si ha una sola variabile di raggruppamento, questo valore è identico alla percentuale della somma totale.

Percentuale di N totale. Percentuale del numero totale di casi in ciascuna categoria.

Percentuale della somma totale..Percentuale della somma totale in ciascuna categoria.

Intervallo. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.

Asimmetria. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con asimmetria positiva ha una coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica uno scostamento dalla simmetria.

deviazione standard. Misura della dispersione intorno alla media. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95% dei casi rientra in due deviazioni standard. Se, ad esempio, l'età media è 45 e la deviazione standard è 10, il 95% dei casi sarà compreso tra 25 e 65 in un distribuzione normale.

Errore standard della curtosi. Il rapporto tra la curtosi e il suo errore standard può essere utilizzato come test di normalità (cioè, si può rifiutare la normalità se il rapporto è inferiore a -2 o superiore a +2). Un valore positivo elevato per la curtosi indica che le code della distribuzione sono più lunghe di quelle di una distribuzione normale; un valore negativo per la curtosi indica code più corte (che diventano simili a quelle di una distribuzione uniforme a forma di scatola).

Errore standard della media. Una misura di quanto il valore della media può variare da campione a campione preso dalla stessa distribuzione. Può essere usata per comparare genericamente la media osservata rispetto a un valore ipotizzato (ovvero, è possibile concludere che i due valori sono diversi se il rapporto della differenza rispetto all'errore standard è inferiore a -2 o maggiore di +2).

Errore standard dell'assimetria. Il rapporto tra la skewness e il suo errore standard può essere utilizzato come test di normalità (cioè, si può rifiutare la normalità se il rapporto è inferiore a -2 o superiore a +2). Un valore positivo elevato per l'asimmetria indicata una coda a destra lunga; un valore negativo estremo indica una coda a sinistra lunga.

Somma. La somma o il totale di tutti i valori non mancanti di tutti i casi.

Varianza. Una misura di dispersione intorno alla media, uguale alla somma dei quadrati delle deviazioni dalla media divisa per un valore inferiore al numero di casi. La varianza è espressa in quadrati dell'unità di misura della variabile.

Cubi OLAP: Differenze

Questa finestra di dialogo consente di calcolare le differenze aritmetiche e percentuali tra variabili di riepilogo o tra gruppi definiti da una variabile di raggruppamento. Le differenze vengono calcolate per tutte le misure selezionate nella finestra di dialogo Cubi OLAP: Statistiche.

Differenze tra variabili. Consente di calcolare le differenze tra coppie di variabili. I valori delle statistiche di riepilogo della seconda variabile di ogni coppia (la variabile sottratta) vengono sottratti dai valori delle statistiche di riepilogo della prima variabile della coppia. Per le differenze percentuali, il valore della variabile di riepilogo della variabile sottratta viene utilizzato al denominatore. È necessario selezionare almeno due variabili di riepilogo nella finestra di dialogo principale prima di specificare le differenze tra variabili.

Differenze tra gruppi di casi. Consente di calcolare le differenze tra coppie di gruppi definiti da una variabile di raggruppamento. I valori delle statistiche di riepilogo della seconda categoria di ogni coppia (la categoria sottratta) vengono sottratti dai valori delle statistiche di riepilogo della prima categoria della coppia. Per le differenze percentuali, il valore delle statistiche di riepilogo della categoria sottratta viene utilizzato al denominatore. È necessario selezionare una o più variabili di raggruppamento nella finestra di dialogo principale prima di specificare le differenze tra gruppi.

Cubi OLAP: Titolo

È possibile modificare il titolo dell'output o aggiungere una didascalia che verrà visualizzata sotto la tabella dell'output. È anche possibile controllare gli a capo automatici nei titoli e nelle didascalie digitando \n dove si desidera inserire un'interruzione della riga nel testo.

Proporzioni

Introduzione delle proporzioni

Proporzioni calcola test e intervalli di confidenza per proporzioni binomiali o differenze di proporzioni. Le statistiche sono disponibili per le proporzioni di un campione (testate contro un valore specificato), campioni accoppiati (variabili diverse) o campioni indipendenti (diversi gruppi di casi). Vengono offerte varie opzioni per i tipi di statistiche di prova e gli intervalli di confidenza. Altre procedure che forniscono alcune delle stesse funzionalità includono **CROSSTABS**, **NPAR TESTS**e **NPTESTS**.

Proporzioni campione singolo
Prove a campione e intervalli di confidenza. L'output include la proporzione osservata, la stima della differenza tra la proporzione di popolazione e la proporzione di popolazione ipotizzata, gli errori standard asintotici sotto ipotesi nulle e alternative, le statistiche di prova specificate con probabilità a due lati, e gli intervalli di confidenza specificati per le proporzioni.

Proporzioni campioni accoppiati
Coppie di campioni accoppiati e intervalli di confidenza per differenze tra le proporzioni. L'output include proporzioni osservate, stime delle differenze di proporzioni della popolazione, errori standard asintotici delle differenze di popolazione sotto ipotesi nulle e alternative, statistiche di prova specificate con probabilità a due lati e intervalli di confidenza specificati per differenze di proporzioni.

Proporzioni campioni indipendenti
Prove di campioni indipendenti e intervalli di confidenza. L'output include proporzioni osservate, stime delle differenze di proporzioni della popolazione, errori standard asintotici delle differenze di popolazione sotto ipotesi nulle e alternative, statistiche di prova specificate con probabilità a due lati e intervalli di confidenza specificati per differenze di proporzioni.

Proporzioni campione singolo

La Procedura One - Sample Proporzioni fornisce test e intervalli di confidenza per singole proporzioni binomiali. I dati sono assunti da un semplice campione casuale e ogni intervallo di prova o di confidenza è un test separato o un intervallo individuale, basato su una proporzione binomiale. L'output include la proporzione osservata, la stima della differenza tra la proporzione di popolazione e la proporzione di

popolazione ipotizzata, gli errori standard asintotici sotto ipotesi nulle e alternative, le statistiche di prova specificate con probabilità a due lati, e gli intervalli di confidenza specificati per le proporzioni.

Esempio

Statistiche

Agresti-Coull, Anscombe, Clopper - Pearson (Esatto), Jeffreys, Logit, Wald, Wald (continuità corretta), Wilson Score, Wilson Score (continuità corretta), Exatto Binomiale, Mid - p Adjusted Binomiale, Score, Score (continuità corretta).

Considerazioni sui dati

Dati

La procedura visualizza le statistiche di prova richieste e le probabilità a due lati, gli intervalli di confidenza per differenze di proporzioni, così come le proporzioni, gli errori standard e conta per ogni gruppo o variabile. La procedura è limitata al massimo di un valore di prova.

Ipotesi

Ottenine One - Sample Proporzioni test

1. Dai menu, scegliere:

   Analizzare > **Confrontare i mezzi** > **A - sample proporzioni ...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili quantitative oggetto del test.
3. Facoltativamente, è possibile:
   - Selezionare le impostazioni dei criteri di successo sotto la sezione **Definisci Successo** :

     Ultimo valore
     Viene utilizzato l'ultimo o il valore più alto tra i valori distinti ordinati nei dati. Questo vale per le variabili numeriche o di stringa. Questa è l'impostazione predefinita.

     Primo valore
     Viene utilizzato il valore primo o più basso tra i valori distinti ordinati nei dati. Questo vale per le variabili numeriche o di stringa.

     Valori
     Uno o più valori specifici paresi. I valori multipli devono essere separati da spazi. Questo vale per le variabili numeriche o di stringa. I valori variabili stringa devono essere racchiusi tra virgolette singole.

     Punto centrale
     Valori a o sopra la metà della gamma dei valori osservati nei dati. Questo vale solo per i dati numerici.

     Punto di divisione
     Valori a o superiore a un valore specificato. Questo vale solo per i dati numerici.
   - Clicca **Intervalli di confidenza ...** per specificare quali tipi di intervalli di confidenza vengono visualizzati o per sopprimere tutti gli intervalli di confidenza.
   - Clicca su **Test ...** per specificare quali tipi di statistiche di prova vengono visualizzate o per sopprimere tutti i test.
   - Clicca su **Valori mancanti ...** per controllare il trattamento dei dati mancanti.
   - Fare clic su **Bootstrap ...** per derivare stime robusta di errori standard e intervalli di confidenza per stime quali la media, la mediana, la proporzione, il rapporto di probabilità, il coefficiente di correlazione o il coefficiente di regressione.
4. Fare clic su **OK**.

Proporzioni One - Sample: Intervalli di confidenza

Dialogo Intervalli di confidenza fornisce opzioni per specificare il livello di copertura e per la selezione dei tipi di intervalli di confidenza visualizzati.

Livello di copertura
Specifica la percentuale di intervallo di confidenza. Deve essere specificato un valore numerico nella gamma (0.100). 95 è l'impostazione predefinita.

Tipi di intervalli
Fornisce opzioni per specificare quali tipi di intervalli di confidenza vengono visualizzati. Le opzioni disponibili includono:

- Agresti-Coull
- Anscombe
- Clopper-Pearson (esatto)
- Jeffreys
- Logit
- Wald
- Wald (continuità corretta)
- Punteggio Wilson
- Punteggio Wilson (continuità corretta)

Specifica Intervalli di confidenza per le Proporzioni di One - Sample

1. Dai menu, scegliere:

   Analizzare > **Confrontare i mezzi** > **A - sample proporzioni ...**
2. Clicca **Intervalli Di Confidenza** per specificare quali tipi di intervalli di confidenza vengono visualizzati o per sopprimere tutti gli intervalli di confidenza.

One - Proporzioni di esempio: Test

Dialogo Test fornisce opzioni per specificare quali tipi di statistiche di prova vengono visualizzate.

Tutti
Tutte le statistiche di prova visualizzano nell'output.

Nessuna
Nessun display di statistiche di prova nell'output.

Binomiale esatto
Visualizza le probabilità binomiali esatte.

Binomiale adattato Mid-p
Visualizza le probabilità binomiali regolate mid - p. Questa è un'impostazione predefinita.

Punteggio
Visualizza la statistica di prova Score Z. Questa è un'impostazione predefinita.

Punteggio (continuità corretta)
Visualizza la statistica di test di punteggio corretto Z.

Wald
Visualizza la statistica del test Wald Z.

Wald (continuità corretta)
Visualizza la statistica del test Wald Z corretto di continuità.

Valore test
Specifica un valore di prova compreso tra 0 e 1. Il valore predefinito è 0.5.

Ottenine One - Sample Proporzioni test

1. Dai menu, scegliere:

   Analizzare > **Confrontare i mezzi** > **A - sample proporzioni ...**
2. Nella finestra di dialogo One - Sample Proporzioni, fare clic su **Test**.
3. Selezionare uno o più dei test disponibili.

One - Proporzioni di esempio: Valori mancanti

La finestra di dialogo Valori mancanti fornisce opzioni per affrontare i valori mancanti.

Ambito dati mancanti

Escludi casi analisi per analisi
Indica l'inclusione di tutti i casi con dati sufficienti sulle variabili utilizzate in ogni particolare analisi. Questa è l'impostazione predefinita.

Escludi casi listwise
Indica l'inclusione di tutti i casi con dati sufficienti su tutte le variabili utilizzate in tutte le analisi.

Valori mancanti definiti dall'utente
Escludere trattando i valori mancanti dell'utente come mancanti. **Includi** ignora le designazioni di valore mancanti dell'utente e tratta i valori mancanti dell'utente come validi.

Definizione delle impostazioni di valore mancanti per le Proporzioni di One - Sample

1. Dai menu, scegliere:

   Analizzare > **Confrontare i mezzi** > **A - sample proporzioni ...**
2. Nella finestra di dialogo One - Sample Proporzioni, clicca su **Valori mancanti**.
3. Selezionare le impostazioni dei valori mancanti desiderati.

Proporzioni campioni accoppiati

La Procedura Pandate - Campioni Proporzioni fornisce test e intervalli di confidenza per la differenza in due proporzioni binomiali correlate o accoppiate. I dati sono assunti da un semplice campione casuale e ogni intervallo di prova o di confidenza è un test separato o un intervallo individuale. L'output include proporzioni osservate, stime delle differenze di proporzioni della popolazione, errori standard asintotici delle differenze di popolazione sotto ipotesi nulle e alternative, statistiche di prova specificate con probabilità a due lati e intervalli di confidenza specificati per differenze di proporzioni.

Esempio

Statistiche
Agresti-Min, Bonett-Price, Newcombe, Wald, Wald (continuità corretta), Binomiale esatta, Binomiale aggiustata a metà, McNemar, McNemar (continuità corretta).

Considerazioni sui dati

Dati

- È richiesto un elenco variabile contenente almeno due variabili.
- Se viene specificato un solo elenco di variabili, ogni membro della lista viene accoppiato con ogni altro membro della lista.

Ipotesi

- Se due elenchi di variabili sono separati da **WITH** senza la parola chiave (**PAIRED**), ogni membro del primo elenco viene associato a ciascun membro del secondo elenco.
- Se due elenchi di variabili sono separati da **WITH** e il secondo elenco è seguito da (**PAIRED**), i membri dei due elenchi in ordine vengono accoppiati: il primo membro del primo elenco viene

accoppiato con il primo membro del secondo elenco, i secondi membri di ciascun elenco vengono accoppiati e così via. Le variabili non corrispondenti vengono ignorate e viene emesso un messaggio di avviso.

Ottenere Pandate - Campioni Proporzioni test

1. Dai menu, scegliere:

   Analizzare > **Confrontare Mezzi** > **Proporzioni - Campioni accoppiati ...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili quantitative oggetto del test.
3. Facoltativamente, è possibile:
   - Selezionare le impostazioni dei criteri di successo sotto la sezione **Definisci Successo** :

     Ultimo valore
     Viene utilizzato l'ultimo o il valore più alto tra i valori distinti ordinati nei dati. Questo vale per le variabili numeriche o di stringa. Questa è l'impostazione predefinita.

     Primo valore
     Viene utilizzato il valore primo o più basso tra i valori distinti ordinati nei dati. Questo vale per le variabili numeriche o di stringa.

     Valori
     Uno o più valori specifici paresi. I valori multipli devono essere separati da spazi. Questo vale per le variabili numeriche o di stringa. I valori variabili stringa devono essere racchiusi tra virgolette singole.

     Punto centrale
     Valori a o sopra la metà della gamma dei valori osservati nei dati. Questo vale solo per i dati numerici.

     Punto di divisione
     Valori a o superiore a un valore specificato. Questo vale solo per i dati numerici.
   - Clicca **Intervalli di confidenza ...** per specificare quali tipi di intervalli di confidenza vengono visualizzati o per sopprimere tutti gli intervalli di confidenza.
   - Clicca su **Test ...** per specificare quali tipi di statistiche di prova vengono visualizzate o per sopprimere tutti i test.
   - Clicca su **Valori mancanti ...** per controllare il trattamento dei dati mancanti.
   - Fare clic su **Bootstrap ...** per derivare stime robusta di errori standard e intervalli di confidenza per stime quali la media, la mediana, la proporzione, il rapporto di probabilità, il coefficiente di correlazione o il coefficiente di regressione.
4. Fare clic su **OK**.

Accoppiata - Proporzioni Campioni: Intervalli di confidenza

Dialogo Intervalli di confidenza fornisce opzioni per specificare il livello di copertura e per la selezione dei tipi di intervalli di confidenza visualizzati.

Livello di copertura
Specifica la percentuale di intervallo di confidenza. Deve essere specificato un valore numerico nella gamma (0.100). 95 è l'impostazione predefinita.

Tipi di intervalli
Fornisce opzioni per specificare quali tipi di intervalli di confidenza vengono visualizzati. Le opzioni disponibili includono:

- Agresti-Min
- Bonett-Price

- Newcombe
- Wald
- Wald (continuità corretta)

Specifica Intervalli di confidenza per Accoppiata - Proporzioni dei campioni

1. Dai menu, scegliere:

   Analizzare > **Confrontare Mezzi** > **Proporzioni - Campioni accoppiati ...**
2. Clicca **Intervalli Di Confidenza** per specificare quali tipi di intervalli di confidenza vengono visualizzati o per sopprimere tutti gli intervalli di confidenza.

Coppie - Campioni Proporzioni: Test

Dialogo Test fornisce opzioni per specificare quali tipi di statistiche di prova vengono visualizzate.

Tutti
Tutte le statistiche di prova visualizzano nell'output.

Nessuna
Nessun display di statistiche di prova nell'output.

Binomiale esatto
Visualizza le probabilità binomiali esatte.

Binomiale adattato Mid-p
Visualizza le probabilità binomiali regolate mid - p. Questa è un'impostazione predefinita.

McNemar
Visualizza la statistica del test Z McNemar . Questa è un'impostazione predefinita.

McNemar (continuità corretta)
Visualizza la statistica del test Z McNemar corretta per la continuità.

Wald
Visualizza la statistica del test Wald Z.

Wald (continuità corretta)
Visualizza la statistica del test Wald Z corretto di continuità.

Ottenere Pandate - Campioni Proporzioni test

1. Dai menu, scegliere:

   Analizzare > **Confrontare Mezzi** > **Proporzioni - Campioni accoppiati ...**
2. Nella finestra di dialogo Pandate - Campioni Proporzioni, clicca su **Test**.
3. Selezionare uno o più dei test disponibili.

Pandate - Proporzioni Campioni: Valori mancanti

La finestra di dialogo Valori mancanti fornisce opzioni per affrontare i valori mancanti.

Ambito dati mancanti

Escludi casi analisi per analisi
Indica l'inclusione di tutti i casi con dati sufficienti sulle variabili utilizzate in ogni particolare analisi. Questa è l'impostazione predefinita.

Escludi casi listwise
Indica l'inclusione di tutti i casi con dati sufficienti su tutte le variabili utilizzate in tutte le analisi.

Valori mancanti definiti dall'utente
Escludere trattando i valori mancanti dell'utente come mancanti. **Includi** ignora le designazioni di valore mancanti dell'utente e tratta i valori mancanti dell'utente come validi.

Definizione delle impostazioni di valore mancanti per Pandate - Campioni Proporzioni

1. Dai menu, scegliere:

   Analizzare > **Confrontare Mezzi** > **Proporzioni - Campioni accoppiati ...**
2. Nella finestra di dialogo Pandate - Campioni Proporzioni, clicca su **Valori mancanti**.
3. Selezionare le impostazioni dei valori mancanti desiderati.

Proporzioni campioni indipendenti

La Procedura Indipendenti - Campioni Proporzioni fornisce test e intervalli di confidenza per la differenza in due proporzioni binomiali indipendenti. I dati sono assunti da un semplice campione casuale e ogni intervallo di prova o di confidenza è un test separato o un intervallo individuale. L'output include proporzioni osservate, stime delle differenze di proporzioni della popolazione, errori standard asintotici delle differenze di popolazione sotto ipotesi nulle e alternative, statistiche di prova specificate con probabilità a due lati e intervalli di confidenza specificati per differenze di proporzioni.

Esempio

Statistiche
Agresti-Min, Bonett-Price, Newcombe, Wald, Wald (continuità corretta), Binomiale esatta, Binomiale aggiustata a metà, McNemar, McNemar (continuità corretta).

Considerazioni sui dati

Dati

- Sono necessarie almeno una variabile dipendente e una singola variabile per identificare i due gruppi da confrontare.
- La variabile di raggruppamento può essere numerica o stringa.

Ipotesi

Ottenere Prove Indipendenti - Campioni Proporzioni

1. Dai menu, scegliere:

   Analizzare > **Confrontare Mezzi** > **Proporzioni dei campioni indipendenti ...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili quantitative oggetto del test.
3. Selezionare un singolo **Raggruppamento Variabile** che identifica i due gruppi da confrontare.
4. Facoltativamente, specificare le impostazioni per la **Variabile di raggruppamento**selezionata.
   - Quando si seleziona **Value (s)** è possibile specificare due valori numerici o di stringa tra parentesi per i valori da confrontare. I valori di stringa devono essere racchiusi tra virgolette singole. I casi con altri valori vengono ignorati.
   - **Midpoint** si applica solo a variabili numeriche. I casi a o superiore al midpoint della distribuzione della variabile di raggruppamento sono assegnati al secondo gruppo, i casi sotto il midpoint sono assegnati al primo gruppo.
   - **Taglia Point** applica solo le variabili numeriche e consente la specifica con parentesi di un unico valore numerico. I casi a o sopra il punto di taglio sulla variabile di raggruppamento sono assegnati al secondo gruppo, i casi sotto il punto di taglio sono assegnati al primo gruppo.
5. Facoltativamente, è possibile:
   - Selezionare le impostazioni dei criteri di successo sotto la sezione **Definisci Successo** :

Ultimo valore
Viene utilizzato l'ultimo o il valore più alto tra i valori distinti ordinati nei dati. Questo vale per le variabili numeriche o di stringa. Questa è l'impostazione predefinita.

Primo valore
Viene utilizzato il valore primo o più basso tra i valori distinti ordinati nei dati. Questo vale per le variabili numeriche o di stringa.

Valori
Uno o più valori specifici paresi. I valori multipli devono essere separati da spazi. Questo vale per le variabili numeriche o di stringa. I valori variabili stringa devono essere racchiusi tra virgolette singole.

Punto centrale
Valori a o sopra la metà della gamma dei valori osservati nei dati. Questo vale solo per i dati numerici.

Punto di divisione
Valori a o superiore a un valore specificato. Questo vale solo per i dati numerici.

- Clicca **Intervalli di confidenza ...** per specificare quali tipi di intervalli di confidenza vengono visualizzati o per sopprimere tutti gli intervalli di confidenza.
- Clicca su **Test ...** per specificare quali tipi di statistiche di prova vengono visualizzate o per sopprimere tutti i test.
- Clicca su **Valori mancanti ...** per controllare il trattamento dei dati mancanti.
- Fare clic su **Bootstrap ...** per derivare stime robusta di errori standard e intervalli di confidenza per stime quali la media, la mediana, la proporzione, il rapporto di probabilità, il coefficiente di correlazione o il coefficiente di regressione.

6. Fare clic su **OK**.

Indipendenti - Proporzioni Campioni: Intervalli di confidenza

Dialogo Intervalli di confidenza fornisce opzioni per specificare il livello di copertura e per la selezione dei tipi di intervalli di confidenza visualizzati.

Livello di copertura
Specifica la percentuale di intervallo di confidenza. Deve essere specificato un valore numerico nella gamma (0.100). 95 è l'impostazione predefinita.

Tipi di intervalli
Fornisce opzioni per specificare quali tipi di intervalli di confidenza vengono visualizzati. Le opzioni disponibili includono:

- Agresti-Caffo
- Brown-Li-Jeffreys
- Hauck-Anderson
- Newcombe
- Newcombe (continuità corretta)
- Wald
- Wald (continuità corretta)

Specifica Intervalli di confidenza per le proporzioni Indipendenti - Campioni

1. Dai menu, scegliere:

   Analizzare > **Confrontare Mezzi** > **Proporzioni dei campioni indipendenti ...**

2. Clicca **Intervalli Di Confidenza** per specificare quali tipi di intervalli di confidenza vengono visualizzati o per sopprimere tutti gli intervalli di confidenza.

Indipendenti - Proporzioni Campioni: Test

Dialogo Test fornisce opzioni per specificare quali tipi di statistiche di prova vengono visualizzate.

Tutti
Tutte le statistiche di prova visualizzano nell'output.

Nessuna
Nessun display di statistiche di prova nell'output.

Hauck-Anderson
Visualizza la statistica di prova Hauck - Anderson Z.

Wald
Visualizza la statistica del test Wald Z.

Wald (continuità corretta)
Visualizza la statistica del test Wald Z corretto di continuità.

Wald H0
Visualizza la statistica del test Wald Z utilizzando le stime della varianza sotto H_0.

Wald H0(continuità corretta)
Visualizza la statistica del test Wald Z corretta per la continuità utilizzando le stime della varianza sotto H_0.

Ottenere Prove Indipendenti - Campioni Proporzioni

1. Dai menu, scegliere:

   Analizzare > **Confrontare Mezzi** > **Proporzioni dei campioni indipendenti ...**
2. Nella finestra di dialogo Indipendente - Campioni Proporzioni, fare clic su **Test**.
3. Selezionare uno o più dei test disponibili.

Independent - Proporzioni Campioni: Valori mancanti

La finestra di dialogo Valori mancanti fornisce opzioni per affrontare i valori mancanti.

Ambito dati mancanti

Escludi casi analisi per analisi
Indica l'inclusione di tutti i casi con dati sufficienti sulle variabili utilizzate in ogni particolare analisi. Questa è l'impostazione predefinita.

Escludi casi listwise
Indica l'inclusione di tutti i casi con dati sufficienti su tutte le variabili utilizzate in tutte le analisi.

Valori mancanti definiti dall'utente
Escludere trattando i valori mancanti dell'utente come mancanti. **Includi** ignora le designazioni di valore mancanti dell'utente e tratta i valori mancanti dell'utente come validi.

Definizione delle impostazioni di valore mancanti per Proporzioni Indipendenti - Campioni

1. Dai menu, scegliere:

   Analizzare > **Confrontare Mezzi** > **Proporzioni dei campioni indipendenti ...**
2. Nella finestra di dialogo Independent - Campioni Proporzioni, clicca su **Valori mancanti**.
3. Selezionare le impostazioni dei valori mancanti desiderati.

Test T

Test T

Sono disponibili tre tipi di test *t*:

Test T: campioni indipendenti (test T: due campioni). Consente di confrontare le medie di una variabile per due gruppi di casi. Vengono fornite statistiche descrittive per ciascun gruppo, il test di Levene di uguaglianza delle varianze, i valori *t* di uguaglianza e non uguaglianza della varianza e un intervallo di confidenza al 95% per la differenza tra le medie.

Test T: campioni accoppiati (test T: dipendente). Consente di confrontare le medie di due variabili per un singolo gruppo. Questo test viene utilizzato anche per disegni di studi di coppie corrispondenti o di controllo casi. Vengono fornite statistiche descrittive per le variabili oggetto del test, la correlazione tra di esse, le statistiche descrittive per le differenze appaiate, il test *t* e un intervallo di confidenza al 95%.

Test T: campione unico. Consente di confrontare la media di una variabile con un valore noto o un valore ipotizzato. Con il test *t* vengono visualizzate anche le statistiche descrittive per le variabili oggetto del test. L'output predefinito include un intervallo di confidenza al 95% per la differenza tra la media della variabile oggetto del test e il valore ipotizzato per il test.

Test T a campioni indipendenti

La procedura di test Independent - Samples T mette a confronto i mezzi per due gruppi di casi e automatizza il calcolo della dimensione dell'effetto di test *t*. Idealmente, per questo test i soggetti dovrebbero essere assegnati in modo casuale a due gruppi, in modo che qualsiasi differenza di risposta sia dovuta al trattamento (o alla mancanza di trattamento) e non ad altri fattori. Ciò non si verifica se si esegue il confronto tra il reddito medio di soggetti maschili e femminili. Non è infatti possibile assegnare in modo casuale una persona al sesso maschile o femminile. In questi casi, è necessario assicurarsi che le differenze relative ad altri fattori non comportino un mascheramento o l'incremento di differenze significative nelle medie. Le differenze nel reddito medio possono essere influenzate da fattori quali il livello di educazione e non solo dal sesso al quale appartengono i soggetti.

Esempio

I pazienti con pressione sanguigna alta vengono assegnati in modo casuale a un gruppo di controllo e a un gruppo di trattamento. Ai soggetti del gruppo di controllo vengono somministrate medicine innocue e ai soggetti del gruppo trattato viene somministrato un nuovo farmaco che si ritiene possa far diminuire la pressione sanguigna. Al termine di un trattamento di due mesi, viene utilizzato il test *t* per due campioni allo scopo di confrontare i valori medi della pressione sanguigna nel gruppo di controllo e nel gruppo di trattamento. La pressione di ogni paziente viene misurata una volta e ciascun paziente appartiene a un solo gruppo.

Statistiche

Per ogni variabile: dimensione del campione, media, deviazione standard, errore standard della media e stima della dimensione dell'effetto per il test *t*. Per la differenza nelle medie: media, errore standard e intervallo di confidenza (è possibile specificare il livello di confidenza). Test: test di Levene per l'eguaglianza delle varianze ed entrambi i test di varianze raggruppate e varianze separate *t* per l'eguaglianza delle medie.

Considerazioni sui dati

Dati

I valori della variabile quantitativa desiderata si trovano in una singola colonna del file di dati. Viene utilizzata una variabile di raggruppamento che include due valori per suddividere i casi in due gruppi. La variabile di raggruppamento può essere numerica (valori quali 1 e 2, o 6,25 e 12,5) oppure una stringa breve (ad esempio *sì* e *no*). In alternativa, è possibile utilizzare una variabile quantitativa, ad esempio *età*, per suddividere i casi in due gruppi specificando un punto di divisione (il punto di divisione 21 suddivide la variabile *età* in un gruppo con meno di 21 anni e in un gruppo con più di 21 anni).

Ipotesi

Per il test *t* di uguaglianza della varianza, le osservazioni dovrebbero essere rappresentate da campioni casuali e indipendenti derivati da distribuzioni normali con la stessa varianza di popolazione. Per il test *t* di inuguaglianza della varianza, le osservazioni dovrebbero essere campioni casuali e indipendenti derivati da distribuzioni normali. Il test *t* per due campioni è sufficientemente robusto per le deviazioni dalla normalità. Durante la verifica grafica delle distribuzioni, controllare che siano simmetriche e che non siano presenti valori anomali.

Ottenimento di un Independent - Samples T Test

1. Dai menu, scegliere:

   Analizza > **Confronta medie** > **Test T: campioni indipendenti...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili quantitative oggetto del test. Per ciascuna variabile viene calcolato un test *t* distinto.
3. Selezionare una variabile di raggruppamento singola e fare clic su **Definisci gruppi** per specificare due codici per i gruppi che si desidera confrontare.
4. Le seguenti impostazioni sono facoltative.
   - Selezionare **Dimensioni degli effetti di stima** per controllare la stima della dimensione dell'effetto di test *t*.
   - Selezionare il **test di omogeneità della varianza** per calcolare la statistica di Levene e verificare l'uguaglianza delle varianze dei gruppi. È possibile ottenere i risultati in una tabella separata.
   - Fare clic su **Opzioni** per controllare il trattamento dei dati mancanti e il livello dell'intervallo di confidenza.
   - Clicca su **Bootstrap** per derivare stime robusta di errori standard e intervalli di confidenza per le stime come la media, la mediana, la proporzione, il rapporto di probabilità, il coefficiente di correlazione o il coefficiente di regressione.

Indipendente - Samples T-Test Definire i gruppi

Per le variabili di raggruppamento numerico, definire i due gruppi per il test *t*specificando due valori o un cutpoint:

- **Usa i valori specificati.** Inserire un valore per il Gruppo 1 e un altro valore per il Gruppo 2. I casi con altri valori sono esclusi dall'analisi. Non è necessario specificare numeri interi (ad esempio, 6,25 e 12,5 sono validi).
- **Punto di divisione.** Immettere un numero che suddivide i valori della variabile di raggruppamento in due insiemi Assegna i casi con valori minori al punto di divisione da un gruppo e i casi con valori maggiori o uguali al punto di divisione dall'altro gruppo.

Per le variabili di raggruppamento delle stringhe, inserire una stringa per il Gruppo 1 e un altro valore per il Gruppo 2, come *sì* e *no*. I casi con altre stringhe sono esclusi dall'analisi.

Test T per campioni indipendenti: Opzioni

Intervallo di confidenza. Per impostazione predefinita, viene visualizzato un intervallo di confidenza del 95% per la differenza fra le medie. Immettere un valore compreso fra 1 e 99 per richiedere un livello di confidenza differente.

Valori mancanti. Se durante un test su più variabili si riscontra in alcune di esse la presenza di dati mancanti, è possibile indicare alla procedura i casi da includere (o da escludere):

- **Escludi casi analisi per analisi.** Per ciascun test *t* vengono utilizzati tutti i casi con dati validi per le variabili verificate. Le dimensioni del campione possono variare in base al test.

- **Escludi casi listwise.** Per ciascun test *t* vengono utilizzati solo i casi con dati validi per tutte le variabili prese in considerazione nei test *t*. La dimensione del campione è costante nei vari test.

Test T a campioni accoppiati

La procedura Test T per campioni appaiati consente di confrontare le medie di due variabili per un singolo gruppo. La procedura calcola le differenze tra i valori delle due variabili per ciascun caso e viene verificato se la media è diversa da 0. La procedura automatizza anche il calcolo della dimensione dell'effetto di test *t*.

Esempio

In uno studio su pazienti con valori elevati della pressione sanguigna, a tutti i pazienti è stata misurata la pressione all'inizio dello studio, è stato somministrato un trattamento e quindi la misurazione è stata ripetuta. Per ciascun soggetto sono quindi disponibili due misurazioni, in genere denominate *precedente* e *successiva*. Un altro disegno per cui viene utilizzato questo test è lo studio di coppie corrispondenti o di controllo casi, in cui ciascun record del file di dati contiene la risposta per il paziente e quella del soggetto di controllo corrispondente. In uno studio sulla pressione sanguigna, è necessario che l'età dei pazienti trattati corrisponda a quella dei controlli (a un paziente di 75 anni deve corrispondere un membro del gruppo di controllo di 75 anni).

Statistiche

Per ogni variabile: media, dimensione del campione, deviazione standard ed errore standard della media. Per ogni coppia di variabili: correlazione, differenza media nei mezzi, test *t* , intervallo di confidenza per la differenza media (è possibile specificare il livello di confidenza), e la stima della dimensione dell'effetto per il test *t*. Deviazione standard ed errore standard della differenza media.

Considerazioni sui dati

Dati

Per ciascun test appaiato, specificare due variabili quantitative (livello di misura in base a intervallo o a rapporto). In uno studio di confronti tra coppie o di casi di controllo, la risposta per ciascun soggetto del test e per il soggetto di controllo corrispondente deve trovarsi nello stesso caso all'interno del file di dati.

Ipotesi

Le osservazioni per ciascuna coppia devono essere effettuate nelle medesime condizioni. Le differenze medie devono essere distribuite normalmente. Le varianze di ciascuna variabile possono essere uguali o non uguali.

Ottenimento di un Pandate - Samples T Test

1. Dai menu, scegliere:

   Analizza > **Confronta medie** > **Test T: campioni accoppiati...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più coppie di variabili.
3. Facoltativamente, modificare/selezionare un'opzione **Stima dimensioni effetto**. Le impostazioni controllano come viene calcolato lo standardizzatore nella stima della correzione di Cohen *d* e Hedges per ogni coppia variabile.

   Deviazione standard della differenza

   Il denominatore utilizzato nella stima della dimensione effetto. *d* di Cohen utilizza la deviazione standard del campione della differenza media. La correzione di Hedges utilizza la deviazione standard del campione della differenza media adattata da un fattore di correzione.

   Deviazione standard della differenza corretta

   Il denominatore utilizzato nella stima della dimensione effetto. *d* di Cohen utilizza la deviazione standard del campione della differenza media adattata dalla correlazione tra le misure. La

correzione di Hedges utilizza la deviazione standard del campione della differenza media adattata dalla correlazione tra le misure, più un fattore di correzione.

Media delle varianze
Il denominatore utilizzato nella stima della dimensione effetto. *d* di Cohen utilizza la radice quadrata della varianza media delle misure. La correzione di Hedges utilizza la radice quadrata della varianza media delle misure, più un fattore di correzione.

4. Facoltativamente, è possibile:
   - Selezionare **Dimensioni degli effetti di stima** per controllare la stima della dimensione dell'effetto di test *t*. Quando l'impostazione viene selezionata, è possibile controllare ulteriormente come viene calcolato lo standardizzatore nella stima della correzione di Cohen *d* e Hedges per ogni coppia variabile.
   - Fare clic su **Opzioni** per controllare il trattamento dei dati mancanti e il livello dell'intervallo di confidenza.
   - Clicca su **Bootstrap** per derivare stime robusta di errori standard e intervalli di confidenza per le stime come la media, la mediana, la proporzione, il rapporto di probabilità, il coefficiente di correlazione o il coefficiente di regressione.

Test T per campioni appaiati: Opzioni

Intervallo di confidenza. Per impostazione predefinita, viene visualizzato un intervallo di confidenza del 95% per la differenza fra le medie. Immettere un valore compreso fra 1 e 99 per richiedere un livello di confidenza differente.

Valori mancanti. Se durante un test su più variabili si riscontra in alcune di esse la presenza di dati mancanti, è possibile indicare alla procedura i casi da includere (o da escludere):

- **Escludi casi analisi per analisi.** Per ciascun test *t* vengono utilizzati tutti i casi con dati validi per la coppia di variabili verificate. Le dimensioni del campione possono variare in base al test.
- **Escludi casi listwise.** Per ciascun test *t* vengono utilizzati solo i casi che includono dati validi per tutte le coppie di variabili verificate. La dimensione del campione è costante nei vari test.

Funzioni aggiuntive di comando t TEST

Il linguaggio della sintassi dei comandi consente inoltre di:

- Effettuare test di T per un campione e per campioni indipendenti tramite un unico comando.
- Confrontare ciascuna variabile con le variabili dell'elenco in test accoppiati (con il sottocomando PAIRS).
- Controllare la stima della dimensione dell'effetto di test *t*(con il comando ES).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test T a campione singolo

La procedura Test T per un campione consente di verificare se la media di una singola variabile è diversa da una costante specificata e automatizza il calcolo della dimensione dell'effetto del test *t*.

Esempi
Un ricercatore può voler testare se il punteggio IQ medio di un gruppo di stundenti è diverso da 100. Oppure un produttore di cereali può prendere un campione di scatole dalla linea di produzione e controllare se il peso medio dei campioni è diverso da 1.3 libre al livello di confidenza del 95%.

Statistiche
Per ogni variabile di prova: media, deviazione standard, errore standard della media e stima della dimensione dell'effetto per il test *t*. La differenza media fra ciascun valore dei dati e il valore oggetto del test ipotizzato, un test *t* che verifica che la differenza sia uguale a 0 e un intervallo di confidenza per la differenza (è possibile specificare il livello di confidenza).

Considerazioni sui dati

Dati

Per confrontare i valori di una variabile quantitativa con un valore oggetto del test ipotizzato, scegliere una variabile quantitativa e immettere un valore oggetto del test ipotizzato.

Ipotesi

In questo test si presume che i dati siano distribuiti in modo normale. Il test è tuttavia sufficientemente robusto per le deviazioni dalla normalità.

Ottenimento di un Test di T - Esempio

1. Dai menu, scegliere:

   Analizza > **Confronta medie** > **Test T: campione unico...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili da confrontare con lo stesso valore ipotizzato.
3. Immettere un valore oggetto del test numerico rispetto al quale viene confrontata ciascuna media del campione.
4. Facoltativamente, è possibile:
   - Selezionare **Dimensioni degli effetti di stima** per controllare la stima della dimensione dell'effetto di test *t*.
   - Fare clic su **Opzioni** per controllare il trattamento dei dati mancanti e il livello dell'intervallo di confidenza.

Test T per un campione: Opzioni

Intervallo di confidenza. Per impostazione predefinita, viene visualizzato un intervallo di confidenza del 95% per la differenza fra la media e il valore oggetto del test ipotizzato. Immettere un valore compreso fra 1 e 99 per richiedere un livello di confidenza differente.

Valori mancanti. Se durante un test su più variabili si riscontra in alcune di esse la presenza di dati mancanti, è possibile indicare alla procedura i casi da includere (o da escludere):

- **Escludi casi analisi per analisi.** Per ciascun test *t* vengono utilizzati tutti i casi con dati validi per le variabili verificate. Le dimensioni del campione possono variare in base al test.
- **Escludi casi listwise.** Per ciascun tes*t* vengono utilizzati solo casi con dati validi per tutte le variabili prese in considerazione nei *t* test richiesti. La dimensione del campione è costante nei vari test.

Funzioni aggiuntive di comando t TEST

Il linguaggio della sintassi dei comandi consente inoltre di:

- Effettuare test di T per un campione e per campioni indipendenti tramite un unico comando.
- Confrontare ciascuna variabile con le variabili dell'elenco in test accoppiati (con il sottocomando PAIRS).
- Controllare la stima della dimensione dell'effetto di test *t*(con il comando ES).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Funzioni aggiuntive di comando t TEST

Il linguaggio della sintassi dei comandi consente inoltre di:

- Effettuare test di T per un campione e per campioni indipendenti tramite un unico comando.
- Confrontare ciascuna variabile con le variabili dell'elenco in test accoppiati (con il sottocomando PAIRS).

- Controllare la stima della dimensione dell'effetto di test *t*(con il comando ES).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

ANOVA a una via

La Procedura One-Way ANOVA produce un'analisi unidiretiva della varianza per una variabile dipendente quantitativa da parte di un singolo fattore (indipendente) e stima la dimensione dell'effetto in un'ANOVA a senso unico. L'analisi della varianza consente di verificare l'ipotesi di uguaglianza di più medie. Questa tecnica è un'estensione del test *t* per due campioni.

Oltre a determinare le differenze tra le medie, è possibile individuare la media che differisce dalle altre. Sono disponibili due tipi di test per il confronto delle medie: contrasti a priori e test post hoc. I contrasti sono test impostati *prima* di eseguire l'esperimento, mentre i test post hoc vengono effettuati *dopo* l'esecuzione dell'esperimento. È inoltre possibile verificare le tendenze presenti tra le categorie.

Esempio
Le ciambelle assorbono quantità variabili di grassi a seconda della modalità di cottura. È stato impostato un esperimento che coinvolge tre tipi di grassi: olio di semi di arachide, olio di mais e strutto. L'olio di semi di arachide e l'olio di mais sono grassi insaturi, mentre lo strutto è un grasso saturo. Oltre a determinare se la quantità di grassi assorbita dipende dal tipo di grasso utilizzato, è possibile impostare un contrasto a priori per determinare se la quantità di grassi assorbita differisce per i grassi saturi e insaturi.

Statistiche
Per ogni gruppo: numero di casi, media, deviazione standard, errore standard della media, minimo, massimo, 95% intervallo di confidenza per la media, e stima della dimensione dell'effetto per in un ANOVA a senso unico. I test di Levene per l'omogeneità di varianza, tabella di analisi della varianza e tests robusti dell'uguaglianza delle medie per ogni variabile dipendente, contrasti a priori specificati dall'utente, test dell'intervallo post hoc e confronti multipli: Bonferroni, Sidak, differenza significativa di Tukey, GT2 di Hochberg, Gabriel, Dunnett, test *F* di Ryan-Einot-Gabriel-Welsch (*F* di R-E-G-W), test dell'intervallo di Ryan-Einot-Gabriel-Welsch (*Q* di R-E-G-W), T2 di Tamhane, T3 di Dunnett, Games-Howell, *C* di Dunnett, test di intervallo multiplo di Duncan, Student-Newman-Keuls (S-N-K), *b* di , Waller-Duncan, Scheffé e differenza meno significativa.

Considerazioni sui dati

Dati
La variabile dipendente dovrebbe essere quantitativa (livello di intervallo di misurazione).

Ipotesi
Ciascun gruppo è un campione casuale indipendente prelevato da una popolazione normale. L'analisi della varianza è uno stimatore robusto degli scostamenti dalla normalità, anche se i dati devono essere simmetrici. I gruppi devono provenire da popolazioni con varianze uguali. Per verificare questa ipotesi, utilizzare il test dell'omogeneità della varianza di Levene.

Ottenimento di un'analisi di One-Way della varianza

1. Dai menu, scegliere:

   Analizza > **Confronta medie** > **ANOVA a una via...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili dipendenti.
3. Selezionare una singola variabile fattore indipendente.

Facoltativamente, è possibile:

- Selezionare **Stima effetto di stima per i test complessivi** per controllare il calcolo della dimensione dell'effetto per il test complessivo. Quando selezionate, i display da tavola "ANOVA Effect Calibro" nell'output.

- Clicca su **Contrasti** per partizionare le somme tra gruppi di quadrati in componenti di tendenza o specificare un contrasto a priori.
- Clicca su **Post Hoc** per utilizzare test di gamma post hoc e confronti multipli pairwise per determinare quali mezzi differiscono.
- Fare clic su **Opzioni** per controllare il trattamento dei dati mancanti e il livello dell'intervallo di confidenza.
- Clicca su **Bootstrap** per derivare stime robusta di errori standard e intervalli di confidenza per le stime come la media, la mediana, la proporzione, il rapporto di probabilità, il coefficiente di correlazione o il coefficiente di regressione.

ANOVA a una via: Contrasti

È possibile suddividere le somme dei quadrati fra gruppi in componenti di tendenza oppure specificare contrasti a priori.

Polinomiale
Consente di suddividere le somme dei quadrati tra gruppi in componenti di tendenza. È possibile verificare una tendenza della variabile dipendente in tutti i livelli ordinati della variabile fattore. Ad esempio, è possibile verificare una tendenza lineare (crescente o decrescente) nei salari in tutti i livelli ordinati del grado di salario più elevato.

- **Grado.** È possibile scegliere un termine polinomiale di ordine 1, 2, 3, 4 o 5.

Coefficienti
Contrasti a priori definiti dall'utente da verificare con la statistica *t*. Specificare un coefficiente per ciascun gruppo (categoria) della variabile fattore e quindi fare clic su **Aggiungi** dopo aver inserito ciascuna voce. I nuovi valori verranno aggiunti alla fine dell'elenco dei coefficienti. Per specificare altri insiemi di contrasti, fare clic su **Successivo**. Utilizzare **Successivo** e **Precedente** per spostarsi tra gli insiemi di contrasti.

Stima della dimensione dell'effetto per i contrasti
Controlla il calcolo della dimensione dell'effetto per la prova complessiva. Quando questa impostazione è abilitata, deve essere selezionata almeno una delle seguenti opzioni per calcolare le dimensioni dell'effetto. Questa impostazione è abilitata quando viene specificato almeno un contrasto e si ottiene una tabella ANOVA Effect Ensioni nell'output.

Utilizza la deviazione standard raggruppata per tutti i gruppi come standardizzatore
Utilizza la deviazione standard raggruppata per tutti i gruppi come standardizzatore nella stima della dimensione dell'effetto. Questa è l'impostazione predefinita ed è disponibile quando viene selezionata **Dimensione di effetto stimata per i contrasti** .

Utilizza la deviazione standard raggruppata per i gruppi coinvolti nel contrasto come standardizzatore
Utilizza la deviazione standard raggruppata per i gruppi coinvolti nel contrasto come standardizzatore. L'impostazione è disponibile quando viene selezionata **Dimensione di effetto stimata per i contrasti** .

L'ordine dei coefficienti è importante in quanto corrisponde all'ordine crescente dei valori delle categorie della variabile fattore. Il primo coefficiente dell'elenco corrisponde al valore di gruppo minimo della variabile fattore e l'ultimo coefficiente corrisponde al valore massimo. Ad esempio, se esistono sei categorie della variabile fattore, i coefficienti –1, 0, 0, 0, 0.5 e 0.5 contrastano il primo gruppo con i gruppi quinto e sesto. Per la maggior parte delle applicazioni, i coefficienti dovrebbero somma a 0. Vengono utilizzati anche i set che non si somma a 0, ma viene visualizzato un messaggio di avviso.

ANOVA a una via: Test Post Hoc

Dopo aver determinato l'esistenza di differenze tra le medie, i test post hoc di intervallo e i confronti pairwise multipli consentono di determinare quale media differisce dalle altre. I test di intervallo multipli consentono di identificare sottoinsiemi omogenei di medie che non differiscono le une dalle altre. Grazie ai confronti pairwise multipli è possibile verificare la differenza tra ciascuna coppia di medie e ottenere

una matrice in cui gli asterischi indicano le medie di gruppo con differenze significative e un livello alfa 0,05.

Presumi varianze uguali

Test Differenza significativa di Tukey, GT2 di Hochberg, Gabriel e Scheffé sono test di confronto multiplo e test di intervallo. Sono disponibili altri test di intervallo, ovvero *b* di Tukey, S-N-K (Student-Newman-Keuls), Duncan, R-E-G-W *F* (test *F* di Ryan-Einot-Gabriel-Welsch), R-E-G-W *Q* (test di intervallo di Ryan-Einot-Gabriel-Welsch) e Waller-Duncan. I test di confronto multiplo disponibili sono Bonferroni, il test Differenza significativa di Tukey, Sidak, Gabriel, Hochberg, Dunnett, Scheffé e LSD (least significant difference).

- *LSD*. Utilizza i test t per eseguire tutti i confronti pairwise tra medie di gruppo. Non viene apportata alcuna correzione al tasso di errore per confronti multipli.
- *Bonferroni*. Utilizza i test *t* per eseguire confronti pairwise tra i mezzi di gruppo, ma controlla la frequenza di errore globale impostando la frequenza di errore per ogni test al tasso di errore sperimentato diviso per il numero totale di test. Il livello di significatività osservato viene pertanto è adattato tenendo conto che si stanno effettuando confronti multipli.
- *Ispezione a sorpresa*. Test per confronti multipli a coppie basato su una statistica t. Sidak corregge il livello di significatività per confronti multipli e fornisce dei limiti più stretti del Bonferroni.
- *Scheffe*. Esegue confronti congiunti a coppie simultanei per tutte le possibili combinazioni di medie a coppie. Usa la distribuzione di campionamento F. Può essere usato per esaminare tutte le possibili combinazioni lineari di medie di gruppo, non solo i confronti pairwise.
- *R - E - G - W F*. La procedura a multipli decrescenti di Ryan-Einot-Gabriel-Welsch basata su un test F.
- *R - E - G - W Q*. La procedura a multipli decrescenti di Ryan-Einot-Gabriel-Welsch basata su un intervallo studentizzato.
- *S - N - K*. Crea tutti i confronti a coppie tra medie utilizzando la distribuzione dell'intervallo studentizzato. Per dimensioni del campione uguali, confronta anche coppie di medie entro sottoinsiemi omogenei, usando una procedura stepwise. Le medie vengono ordinate dalla più alta alla più bassa e vengono verificate per prime le differenze estreme.
- *Tukey*. Utilizza la statistica dell'intervallo studentizzato per effettuare tutti i confronti pairwise tra gruppi. Imposta il tasso di errore sperimentale al valore del tasso di errore per l'insieme di tutti i confronti pairwise.
- *B di Tukey*. Utilizza la distribuzione dell'intervallo studentizzato per effettuare confronti a coppie tra gruppi. Il valore critico è la media fra il corrispondente valore per il test HSD (honestly significant difference) di Tukey e quello di Student-Newman-Keuls.
- *Duncan*. Crea confronti pairwise utilizzando un ordine stepwise di confronti identico all'ordine utilizzato dal test di Student - Newman - Keuls, ma imposta un livello di protezione per il tasso di errore per la raccolta dei test, piuttosto che un tasso di errore per i singoli test. Usa la statistica dell'intervallo studentizzato.
- *GT2 di Hochberg*.Confronto multiplo e test di intervallo che utilizza il modulo massimo studentizzato. È simile al test HSD (honestly significant difference) di Tukey.
- *Gabriel*. Test per confronti a coppie che ha utilizzato il modulo massimo studentizzato ed è generalmente più potente del GT2 di Hochberg quando le dimensioni delle celle non sono uguali. Se la variabilità delle dimensioni delle celle risulta molto alta, il test di Gabriel può diventare poco conservativo.
- *Waller - Duncan*. Test per confronti multipli basato su una statistica t. Utilizza un approccio bayesiano.
- *Dunnett*. Pairwise multiplo confronto *t* test che confronta una serie di trattamenti contro una singola media di controllo. L'ultima categoria è la categoria di controllo predefinita. In alternativa, è possibile scegliere la prima categoria. **2 - sided** test che la media a qualsiasi livello (tranne la categoria di controllo) del fattore non è uguale a quella della categoria di controllo. **< Control** test se la media a qualsiasi livello del fattore è inferiore a quella della categoria di controllo. **> Controllo** test se la media a qualsiasi livello del fattore è maggiore di quella della categoria di controllo.

Non presumere varianze uguali

I test di confronto multiplo che non presuppongono l'uguaglianza delle varianze sono: T2 di Tamhane, T3 di Dunnett, Games-Howell e *C* di Dunnett.

- *Tamhane's T2*. Test per confronti pairwise conservativi basato su un test t. Questo test è appropriato quando le varianze non sono uguali.
- *Dunnett's T3*. Test per confronti a coppie basato sul modulo massimo studentizzato. Questo test è appropriato quando le varianze non sono uguali.
- *Games-Howell*. Test per confronti pairwise che a volte è liberale. Questo test è appropriato quando le varianze non sono uguali.
- *Dunnett's C*. Test per confronti a coppie basato sull'intervallo studentizzato. Questo test è appropriato quando le varianze non sono uguali.

Nota: Potrebbe essere più facile interpretare l'output da test post hoc se si deseleziona **Nascondi righe e colonne vuote** nella finestra di dialogo Proprietà della tabella (in una tabella pivot attivata, scegliere **Proprietà Tabella** dal menu Formato).

Test sull'ipotesi null

Specifica come viene gestito il livello di significatività (alfa) per il test post hoc.

Utilizzare lo stesso livello di significatività (alpha) come impostazioni nelle Opzioni
Quando selezionato, utilizza la stessa impostazione che viene specificata nella finestra di dialogo Opzioni.

Specifica il livello di significatività (alfa) per il test post hoc
Quando selezionato, è possibile specificare il livello di significatività (alpha) nel campo **Livello** .

Come ottenere test post hoc per l'ANOVA a una via

ANOVA a una via: Opzioni

Statistiche
Consente di scegliere una o più delle seguenti opzioni:

Descrittive
Consente di calcolare il numero di casi, la media, la deviazione standard, l'errore standard della media, il valore minimo e massimo e gli intervalli di confidenza al 95% per ciascuna variabile dipendente di ciascun gruppo.

Effetti fissi e casuali
Consente di visualizzare la deviazione standard, l'errore standard e l'intervallo di confidenza del 95% per il modello degli effetti fissi e l'errore standard, l'intervallo di confidenza del 95% e la stima della varianza tra componenti per il modello degli effetti casuali.

Test di omogeneità della varianza
Consente di calcolare il test di Levene per verificare l'uguaglianza tra gruppi di variabili. Questo test non si basa sull'ipotesi di normalità.

Brown-Forsythe
Calcola la statistica Brown - Forsythe per testare l'uguaglianza dei mezzi di gruppo. Questa statistica è preferibile alla statistica *F* nel caso in cui non sia valida l'ipotesi di uguaglianza della varianza.

Welch
Calcola la statistica Welch per testare l'uguaglianza dei mezzi di gruppo. Questa statistica è preferibile alla statistica *F* nel caso in cui non sia valida l'ipotesi di uguaglianza della varianza.

Valori mancanti
Consente di controllare la modalità di elaborazione dei valori mancanti.

Escludi casi analisi per analisi
Un caso con un valore mancante per la variabile dipendente o il fattore variabile per una determinata analisi non viene utilizzato in tale analisi. Non verranno utilizzati nemmeno i casi che non rientrano nell'intervallo specificato per la variabile fattore.

Escludi casi listwise
Sono esclusi da tutte le analisi i casi con valori mancanti per la variabile fattore o per qualsiasi variabile dipendente inclusa nell'elenco dipendente nella finestra di dialogo principale. Se non sono state specificate più variabili dipendenti, l'opzione non produrrà alcun effetto.

Intervallo di confidenza
Per impostazione predefinita, viene visualizzato un intervallo di confidenza del 95% per la differenza fra la media e il valore oggetto del test ipotizzato. Immettere un valore compreso fra 1 e 99 per richiedere un livello di confidenza differente.

Grafico delle medie
Visualizza un grafico che compone i mezzi del sottogruppo (i mezzi per ogni gruppo definito dai valori della variabile fattore).

Per specificare opzioni per l'ANOVA a una via

Funzioni aggiuntive del comando ONEWAY

Il linguaggio della sintassi dei comandi consente inoltre di:

- Ottenere statistiche con effetti fissi e casuali. Deviazione standard, errore standard della media e intervalli di confidenza al 95% per il modello con effetti fissi. Errore standard, intervalli di confidenza al 95% e stima della varianza dei componenti per il modello con effetti casuali (con il sottocomando `STATISTICS=EFFECTS`).
- Specificare i livelli alfa per la differenza meno significativa, i test di confronto multiplo di Bonferroni, Duncan e Scheffé (con il sottocomando `RANGES`).
- Scrivere una matrice di medie, deviazioni standard e frequenze, oppure leggere una matrice di medie, frequenze, varianze raggruppate e gradi di libertà per le varianze raggruppate. Queste matrici possono essere usate al posto dei dati grezzi per effettuare l'analisi della varianza a una via (con il sottocomando `MATRIX`).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Analisi GLM univariato

La procedura GLM univariato consente di eseguire un'analisi di regressione e un'analisi della varianza per una variabile dipendente tramite uno o più fattori e/o variabili. Le variabili fattore suddividono la popolazione in gruppi. Con questa procedura GLM (modello lineare generalizzato, General Linear Model) è possibile verificare ipotesi nulla relative agli effetti di altre variabili sulle medie di vari raggruppamenti di una sola variabile dipendente. È possibile analizzare le interazioni tra fattori e gli effetti di singoli fattori, alcuni dei quali possono essere casuali. È inoltre possibile includere gli effetti delle covariate e le interazioni tra covariate e fattori. Nell'analisi di regressione, le variabili indipendenti (stimatori) vengono specificate come covariate.

È possibile verificare sia modelli bilanciati che modelli non bilanciati. Un disegno è bilanciato se ciascuna cella del modello include lo stesso numero di casi. Oltre alla verifica delle ipotesi, la procedura GLM univariato consente di ottenere stime dei parametri.

Per la verifica di ipotesi sono disponibili contrasti a priori usati di frequente. Dopo che da un test *F* globale è risultata una certa significatività, è inoltre possibile eseguire test post hoc per valutare le differenze tra medie specifiche. L'opzione Medie marginali stimate consente di ottenere stime dei valori medi previsti delle celle incluse nel modello. I grafici di profilo, o grafici di interazione, di tali medie consentono di visualizzare in modo semplice alcune delle relazioni.

Residui, valori attesi, distanza di Cook e valori di leva possono essere salvati come variabili nel file di dati per la verifica di ipotesi.

Minimi quadrati pesati consente di specificare una variabile per l'assegnazione di pesi diversi alle osservazioni per un'analisi di minimi quadrati pesati (WLS), in alcuni casi per compensare la diversa precisione della misura.

Esempio. Vengono raccolti i dati per i singoli partecipanti della maratona di Chicago per diversi anni. Il tempo impiegato da ciascun partecipante per completare la maratona è la variabile dipendente. Altri fattori presi in considerazione sono le condizioni metereologiche (freddo, caldo o temperatura moderata), il numero di mesi di allenamento, il numero di maratone corse in precedenza e il sesso. L'età è considerata una covariata. Dallo studio può risultare che il sesso rappresenta un effetto significativo, così come l'interazione tra sesso e condizioni meteorologiche.

Metodi. Le somme dei quadrati di tipo I, tipo II, tipo III e tipo IV possono essere utilizzate per valutare ipotesi differenti. Il metodo predefinito è il Tipo III.

Statistiche. Test post hoc di intervallo e confronti multipli: differenze meno significative, Bonferroni, Sidak, Scheffé, *F* multiplo di Ryan-Einot-Gabriel-Welsch, intervallo multiplo di Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls, differenza significativa di Tukey, *b* di Tukey, Duncan, GT2 di Hochberg, Gabriel, test *t* di Waller-Duncan, Dunnett (a 1 via e a 2 vie), T2 di Tamhane, T3 di Dunnett, Games-Howell e *C* di Dunnett. Statistiche descrittive: medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti in tutte le celle. Test di Levene per l'omogeneità della varianza.

Grafici. Di confronto tra diffusione e livello, residuo e di profilo (interazione).

Considerazioni sui dati di GLM univariato

Dati. La variabile dipendente è quantitativa. I fattori sono categoriali. Vi possono essere associati valori numerici o valori stringa composti da un massimo di otto caratteri. Le covariate sono variabili quantitative correlate alla variabile dipendente.

Ipotesi. I dati sono costituiti da un campione casuale derivato da una popolazione normale in cui tutte le varianze di cella sono uguali. L'analisi della varianza è uno stimatore robusto degli scostamenti dalla normalità, anche se i dati devono essere simmetrici. Per la verifica di ipotesi, è possibile usare test di omogeneità della varianza e i grafici di diffusione vs. intensità. È inoltre possibile esaminare residui e grafici dei residui.

Per ottenere tabelle di GLM univariato

1. Dai menu, scegliere:

   Analizza > **Modello lineare generale** > **Univariata...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una variabile dipendente.
3. Selezionare le variabili per l'opzione Fattori fissi, Fattori casuali o Covariate, a seconda dei dati in uso.
4. È inoltre possibile utilizzare l'opzione Peso WLS per specificare una variabile peso per l'analisi dei minimi quadrati pesati. Casi con valore 0, negativo o mancante per la variabile peso saranno esclusi dall'analisi. Non è possibile utilizzare come variabile peso una variabile già inclusa nel modello.

GLM – Univariato: Modello

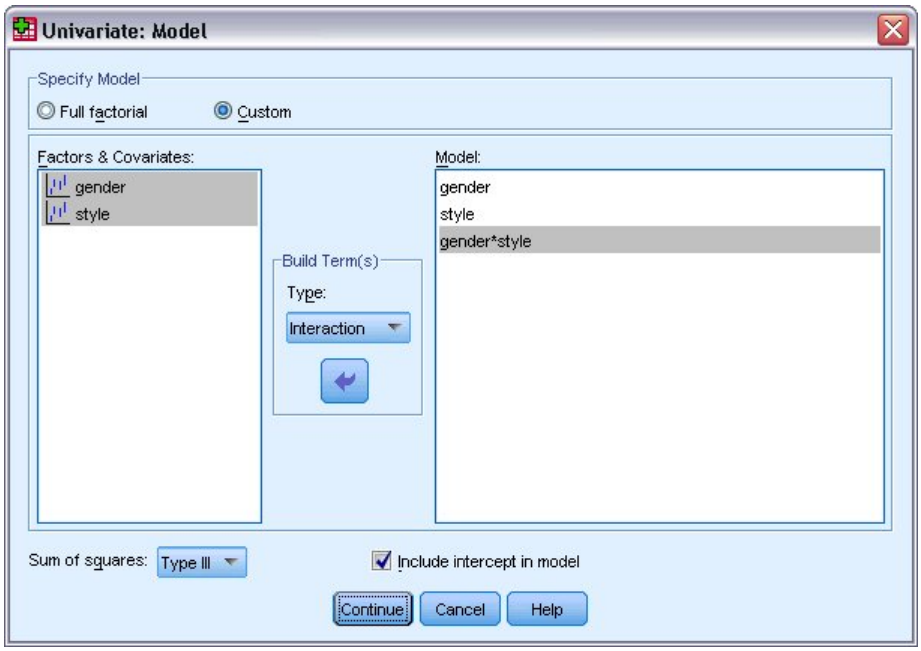


Figura 1. Finestra di dialogo Univariata: Modello

Specifica modello. Un modello fattoriale completo contiene tutti gli effetti principali dei fattori e delle covariate e tutte le interazioni fattore per fattore. Non contiene interazioni di covariate Selezionare **Personalizzato** per specificare un solo sottoinsieme di interazioni o interazioni dei fattori e covariate. È necessario indicare tutti i termini da includere nel modello.

Fattori e covariate. I fattori e le covariate sono elencati.

Modello. Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato **Personalizzato**, è possibile selezionare gli effetti principali e le interazioni desiderate per l'analisi da eseguire.

Somma dei quadrati. Metodo per il calcolo della somma dei quadrati. Il metodo della somma dei quadrati in genere utilizzato con modelli bilanciati o non bilanciati privi di celle mancanti è il Tipo III.

Includi intercetta nel modello. L'intercetta viene in genere inclusa nel modello. Se è possibile presumere che i dati passino attraverso l'origine, l'intercetta può essere esclusa.

Creazione di termini e termini personalizzati

Crea termini
Utilizzare questa opzione quando si desidera includere termini non nidificati di un certo tipo (ad esempio, effetti principali) per tutte le combinazioni di una serie selezionata di fattori e covariate.

Crea termini personalizzati
Utilizzare questa opzione quando si desidera includere termini nidificati o creare esplicitamente una variabile termine dalla variabile. La creazione di un termine nidificato richiede le seguenti operazioni:

Somma dei quadrati

Per il modello è possibile scegliere un tipo di somma dei quadrati. Il Tipo III, il tipo predefinito, è quello usato più di frequente.

Tipo I. Questo metodo è noto anche come decomposizione gerarchica del metodo Somma dei quadrati. Ciascun termine viene adattato solo per i termini del modello che lo precedono. Il metodo Somma dei quadrati Tipo I è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato in cui gli effetti principali vengono specificati prima degli effetti di interazione di ordine 1, ciascuno dei quali viene a sua volta specificato prima degli effetti di interazione di ordine 2 e così via.
- Un modello di regressione polinomiale in cui qualsiasi termine di ordine più basso è specificato prima dei termini di ordine più elevato.
- Un modello nidificato in modo puro in cui il primo effetto specificato è nidificato nel secondo, il quale è a sua volta nidificato nel terzo e così via. Questo tipo di nidificazione può essere specificato esclusivamente tramite la sintassi.

Tipo II. Questo metodo consente di calcolare le somme dei quadrati di un effetto del modello adattato per tutti gli altri effetti "appropriati". È considerato appropriato un effetto corrispondente a tutti gli effetti che non includono l'effetto in esame. Il metodo della somma dei quadrati Tipo II è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato.
- Qualsiasi modello che include solo effetti principali del fattore.
- Qualsiasi modello di regressione.
- Un disegno nidificato in modo puro. Questo tipo di nidificazione può essere specificato tramite la sintassi.

Tipo III. Tipo predefinito. Questo metodo calcola le somme dei quadrati di un effetto nel disegno come la somma dei quadrati adattata per qualsiasi altro effetto che non lo contiene e ortogonale rispetto agli eventuali effetti che lo contengono. Il vantaggio associato a questo tipo di somme dei quadrati è che non varia al variare delle frequenze di cella, a condizione che la forma generale di stimabilità rimanga costante. È pertanto considerato utile per modelli non bilanciati privi di celle mancanti. In un disegno fattoriale privo di celle mancanti, questo metodo equivale alla tecnica dei quadrati delle medie pesate di Yates. Il metodo della somma dei quadrati Tipo III è in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I e il Tipo II.
- Qualsiasi modello bilanciato o non bilanciato e privo di celle vuote.

Tipo IV. Questo metodo è specifico per situazioni con celle mancanti. Per qualsiasi effetto *F* nel disegno, se *F* non è incluso in nessun altro effetto, allora Tipo IV = Tipo III = Tipo II. Se invece *F* è incluso in altri effetti, con il Tipo IV i contrasti creati tra i parametri in *F* vengono distribuiti equamente tra tutti gli effetti di livello superiore. Il metodo della somma dei quadrati Tipo IV viene in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I e il Tipo II.
- Qualsiasi modello bilanciato e non bilanciato contenente celle vuote.

GLM – Univariato: Contrasti

I contrasti consentono di verificare il grado di differenza tra i livelli di un fattore. È possibile specificare un contrasto per ciascun fattore del modello (in un modello a misure ripetute, un contrasto per ciascun fattore tra soggetti). I contrasti rappresentano combinazioni lineari dei parametri.

GLM univariato. Il test sull'ipotesi è basato sull'ipotesi nulla **LB** = 0, dove **L** è la matrice dei coefficienti di contrasto e **B** è il vettore dei parametri. Quando si specifica un contrasto, viene creata una matrice **L**. Le colonne della matrice **L** corrispondenti al fattore corrispondono al contrasto. Le altre colonne vengono adattate in modo che la matrice **L** possa essere stimata.

L'output include una statistica *F* per ciascun insieme di contrasti. Per le differenze dei contrasti vengono inoltre visualizzati gli intervalli di confidenza simultanei di tipo Bonferroni basati su una distribuzione *t* di Student.

Contrasti disponibili

Sono disponibili i contrasti deviazione, semplici, differenza, Helmert, ripetuti e polinomiali. Per i contrasti deviazione e i contrasti semplici, è possibile stabilire se la categoria di riferimento corrisponde alla prima o all'ultima categoria.

Tipi di contrasto

Deviazione. Consente di confrontare la media di ciascun livello, a eccezione di una categoria di riferimento, con la media di tutti i livelli (media principale). L'ordine dei livelli dei fattori può essere un ordine qualsiasi.

Semplice. Consente di confrontare la media di ciascun livello con la media di un livello specifico. Questo tipo di contrasto risulta utile quando è disponibile un gruppo di controllo. Come categoria di riferimento, è possibile scegliere la prima o l'ultima categoria.

Differenza. Consente di confrontare la media di ciascun livello (a eccezione del primo) con la media dei livelli precedenti. Questo tipo di contrasto è a volte definito contrasto inverso di Helmert.

Helmert. Consente di confrontare la media di ciascun livello del fattore (a eccezione dell'ultimo) con la media dei livelli successivi.

Ripetuto. Consente di confrontare la media di ciascun livello (a eccezione dell'ultimo) con la media del livello successivo.

Polinomiale. Consente di confrontare l'effetto lineare, quadratico, cubico e così via. Tutte le categorie del primo grado di libertà includono l'effetto lineare, quelle del secondo includono l'effetto quadratico e così via. Questi contrasti sono spesso usati per la stima delle tendenze polinomiali.

GLM – Univariato: Grafici di profilo

I grafici di profilo, o grafici di interazione, risultano utili per il confronto delle medie marginali di un modello. Un grafico di profilo è un grafico a linee in cui ciascun punto indica la media marginale stimata di una variabile dipendente (adattata per le covariate) in corrispondenza di un solo livello di un fattore. È possibile utilizzare i livelli di un secondo fattore per creare linee distinte. È possibile utilizzare ciascun livello di un terzo fattore per creare un grafico distinto. Tutti gli eventuali fattori casuali e fissi sono disponibili per i grafici. In analisi multivariate i grafici di profilo vengono creati per ciascuna variabile dipendente. Nei grafici di profilo per un'analisi a misure ripetute è possibile includere sia fattori entro soggetti che fattori tra soggetti. GLM Multivariato e GLM Misure ripetute sono disponibili solo se è stato installato il modulo Advanced Statistics.

Il grafico di profilo di un fattore mostra se le medie marginali stimate aumentano o diminuiscono tra i vari livelli. Nel caso di due o più fattori, le linee parallele indicano che tra i fattori non esiste alcuna interazione, ovvero che è possibile analizzare i livelli di un solo fattore. Le linee che si incrociano indicano invece che esiste un'interazione.

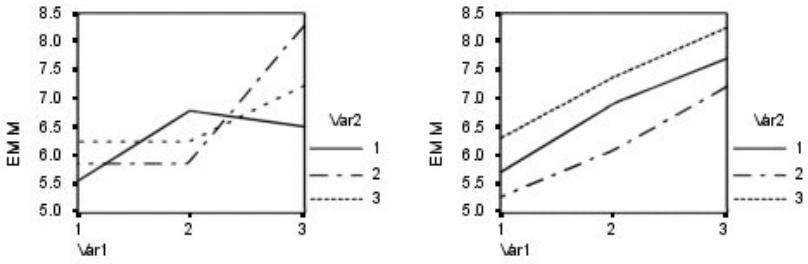


Figura 2. Grafico con linee non parallele (a sinistra) e grafico con linee parallele (a destra)

I grafici definiti tramite la selezione dei fattori per l'asse orizzontale e, se lo si desidera, dei fattori di linee e di grafici separati devono essere inclusi nell'elenco dei grafici.

Tipo di diagramma
Il grafico può essere a linee o a barre.

Barre degli errori
È possibile includere barre degli errori che rappresentano l'intervallo di confidenza o un numero di errori standard. L'intervallo di confidenza si basa sul livello di significatività specificato nella finestra di dialogo Opzioni.

Includi riga di riferimento per la media principale
Include una riga di riferimento che rappresenta la media principale generale.

Asse Y inizia a 0
Per i grafici di linea con tutti i valori positivi o tutti negativi, costringe l'asse Y a partire a 0. I grafici a barre iniziano sempre (o includono) 0.

GLM – Univariato: Opzioni

In questa finestra di dialogo sono disponibili statistiche opzionali. Le statistiche vengono calcolate tramite un modello di effetti fissi.

Visualizzazione. Selezionare **Statistica descrittiva** per produrre medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle. La funzione **Stima della dimensione degli effetti** fornisce un valore eta al quadrato parziale per ciascun effetto e per ciascuna stima del parametro. La statistica eta al quadrato consente di ottenere la proporzione della variabilità totale attribuibile a un fattore. Selezionare **Potenza osservata** per ottenere la potenza del test nel caso in cui l'ipotesi alternativa sia basata sul valore osservato. Selezionare **Stime del parametro** per ottenere stime del parametro, errori standard, test *t*, intervalli di confidenza e la potenza osservata per ciascun test. Selezionare **Matrice dei coefficienti di contrasto** per ottenere la matrice **L**.

La funzione **Test di omogeneità** produce i test di Levene per l'omogeneità della varianza per ogni variabile dipendente su tutte le combinazioni di livello dei fattori tra soggetti, solo per i fattori tra soggetti. Le opzioni Grafici di diffusione vs. densità e Grafici dei residui risultano utili per la verifica di ipotesi sui dati. Se non è disponibile alcun fattore, questa opzione risulta disattivata. Selezionare **Grafici dei residui** per ottenere un grafico dei residui osservati, attesi e standardizzati per ciascuna variabile dipendente. Questi grafici risultano utili per l'analisi dell'ipotesi di uguaglianza della varianza. Selezionare **Mancanza di adattamento** per controllare se la relazione tra la variabile dipendente e le variabili indipendenti può essere descritta in modo adeguato dal modello. **Funzione Generale stimabile (s)** consente di costruire test di ipotesi personalizzati in base alle funzioni generali stimabili. Le righe in qualsiasi matrice di coefficiente di contrasto sono combinazioni lineari della funzione generale stimabile (s).

I **Test dell'eteroschedasticità** sono disponibili per verificare se la varianza di errore (per ogni variabile dipendente) dipende dai valori delle variabili indipendenti. Per il **Test di Breusch-Pagan**, **Test di Breusch-Pagan modificato** e **Test F** è possibile specificare il modello su cui si basa il test. Per impostazione predefinita, il modello consiste in un termine costante, un termine che è lineare nei valori previsti, un termine che è quadratico nei valori previsti e un termine di errore.

Stime parametro con errori standard robusti visualizza una tabella di stime parametro insieme a errori standard robusti o di congruenza con l'eteroschedasticità, e le statistiche *t*, i valori di significatività e gli intervalli di confidenza utilizzati per gli errori standard robusti. Sono disponibili cinque diversi metodi per la stima della matrice di covarianza robusta.

HC0
In base al campione ampio o asintotico originale, lo stimatore robusto, empirico o "sandwich" della matrice di covarianza delle stime parametro. La parte intermedia del sandwich contiene residui di minimi quadrati ordinari o pesati.

HC1
Una modifica di campione finito HC0, moltiplicato per N/(N-p), dove N è la dimensione del campione e p è il numero dei parametri non ridondanti nel modello.

HC2
Una modifica HC0 che coinvolge la divisione del residuo quadrato per 1-h, dove h è la leva per il caso.

HC3
Una modifica HC0 che si approssima a uno stimatore jackknife. I residui quadrati sono divisi per il quadrato di 1-h.

HC4
Una modifica HC0 che divide i residui quadrati per 1-h a una potenza che varia in base a h, N e p, con un limite superiore di 4.

Livello di significatività. Potrebbe risultare utile adattare il livello di significatività usato nei test post hoc e il livello di confidenza usato per la costruzione degli intervalli di confidenza. Il valore specificato viene inoltre usato per il calcolo della potenza osservata per il test. Quando si specifica un livello di significatività, nella finestra di dialogo viene visualizzato il livello di intervalli di confidenza associato.

Funzioni aggiuntive del comando UNIANOVA

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare gli effetti nidificati nel disegno (tramite il sottocomando DESIGN).
- Specificare test di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando TEST).
- Specificare contrasti multipli (tramite il sottocomando CONTRAST).
- Includere valori mancanti definiti dall'utente (tramite il sottocomando MISSING).
- Specificare criteri EPS (tramite il sottocomando CRITERIA).
- Creare una matrice **L**, una matrice **M** o una matrice **K** personalizzata (utilizzando i sottocomandi LMATRIX, MMATRIX e KMATRIX).
- Per i contrasti deviazione e i contrasti semplici, specificare una categoria di riferimento intermedia (tramite il sottocomando CONTRAST).
- Specificare metrica per contrasti polinomiali (tramite il sottocomando CONTRAST).
- Specificare termini di errore per confronti post-hoc (tramite il sottocomando POSTHOC).
- Calcolare medie marginali stimate per qualsiasi fattore o interazione tra fattori nell'elenco dei fattori (tramite il sottocomando EMMEANS).
- Assegnare un nome alle variabili temporanee (tramite il sottocomando SAVE).
- Costruire un file di dati matrice di correlazione (tramite il sottocomando OUTFILE).
- Costruire un file di dati matrice contenente statistiche derivate dai dati della tabella ANOVA tra soggetti (tramite il sottocomando OUTFILE).
- Salvare la matrice del disegno in un nuovo file di dati (tramite il sottocomando OUTFILE).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

GLM: Confronti post hoc

Test di confronti multipli post hoc. Dopo aver determinato l'esistenza di differenze tra le medie, i test post hoc di intervallo e i confronti pairwise multipli consentono di determinare quale media differisce dalle altre. I confronti vengono eseguiti su valori a cui non è stato apportato alcun adattamento. Questi test vengono usati solo con fattori fissi tra soggetti. Nella procedura GLM a misure ripetute, questi test non sono disponibili se non sono presenti fattori tra soggetti e i test di confronto multiplo post hoc vengono eseguiti per la media tra i livelli dei fattori entro soggetti. Per la procedura GLM multivariato, i test post hoc vengono eseguiti separatamente per ciascuna variabile dipendente. GLM Multivariato e GLM Misure ripetute sono disponibili solo se è stato installato il modulo Advanced Statistics.

I test di confronto multiplo usati più di frequente sono il test di Bonferroni e il test HSD di Tukey. Il **test di Bonferroni**, basato sulla statistica *t* di Student, consente di adattare il livello di significatività osservato in base al fatto che vengono eseguiti confronti multipli. Il **test t di Sidak** adatta inoltre il livello di significatività ed è più restrittivo del test di Bonferroni. Il **test HSD di Tukey** utilizza la statistica di intervallo studentizzato per effettuare tutti i confronti pairwise tra gruppi e imposta il tasso di errore sperimentale sul valore del tasso di errore per l'insieme di tutti i confronti pairwise. Quando si eseguono test su un elevato numero di coppie di medie, il test HSD di Tukey risulta più efficace rispetto al test di Bonferroni. Nel caso di un numero limitato di coppie, risulta invece più efficace il test di Bonferroni.

Il test di **Hochberg (GT2)** è simile al test HSD di Tukey, ma utilizza il modulo massimo studentizzato. Il test di Tukey risulta in genere più efficace. Anche il **test per confronti pairwise di Gabriel** utilizza il modulo massimo studentizzato ed è in genere più indicativo del test di Hochberg (GT2) quando le

dimensioni delle celle sono diverse. Se la variabilità delle dimensioni delle celle risulta molto alta, il test di Gabriel può diventare poco conservativo.

Il **test t per confronti multipli pairwise di Dunnett** confronta un insieme di trattamenti con una media di controllo singola. L'ultima categoria è la categoria di controllo predefinita. In alternativa, è possibile scegliere la prima categoria. È inoltre possibile scegliere un test a 2 vie oppure a 1 via. Per verificare che la media in qualsiasi livello dei fattori (a eccezione della categoria di controllo) non sia uguale a quella della categoria di controllo, è necessario usare un test a due sensi. Per verificare se la media a qualsiasi livello del fattore è inferiore a quella della categoria di controllo, selezionare **< Control**. Allo stesso modo, per verificare se la media a qualsiasi livello del fattore è superiore a quella della categoria di controllo, selezionare **> Control**.

Ryan, Einot, Gabriel e Welsch (R-E-G-W) hanno sviluppato due test di intervallo decrescenti multipli. Le procedure a multipli decrescenti verificano in primo luogo se tutte le medie sono uguali. Se le medie non risultano tutte uguali, il test di uguaglianza viene eseguito su un sottoinsieme di medie. Il test **R-E-G-W F** è basato su un test *F*, mentre **R-E-G-W Q** è basato sull'intervallo studentizzato. Questi test risultano più efficaci rispetto ai test di intervallo multiplo di Duncan e Student-Newman-Keuls, che sono pure procedure a intervalli decrescenti multipli. È tuttavia consigliabile non usarli con celle di dimensioni non uguali.

Quando le varianze non sono uguali, utilizzare **T2 di Tamhane** (test per confronti pairwise conservativo basato su un test *t*), **T3 di Dunnett** (test per confronti pairwise basato sui moduli massimi studentizzati), **il test per confronti pairwise di Games-Howell** (a volte liberale) o **C di Dunnett** (test per confronti pairwise basato sull'intervallo studentizzato). Notare che questi test non sono validi e non verranno eseguiti se il modello contiene più fattori.

Il test di intervallo multiplo di Duncan, il test di Student-Newman-Keuls (**S-N-K**) e il test **b di Tukey** sono test di intervallo che classificano le medie raggruppate e calcolano un valore di intervallo. Questi test sono usati meno frequentemente dei test descritti in precedenza.

Il **test t di Waller-Duncan** utilizza un approccio bayesiano. Si tratta di un test di intervallo che usa la media armonica della dimensione del campione nel caso di dimensioni del campione non uguali.

Il livello di significatività del test di **Scheffé** è progettato per consentire il test di tutte le possibili combinazioni lineari delle medie di gruppo, non solo dei confronti pairwise disponibili in questa funzione. Ne risulta che il test di Scheffé è spesso più conservativo di altri test, ciò significa che è richiesta una maggiore differenza tra le medie per la significatività.

Il test di confronto multiplo pairwise Differenza meno significativa o **LSD**, è equivalente a più test *t* tra tutte le coppie di gruppi. Lo svantaggio di questo test è che non viene eseguito alcun tentativo di adattamento del livello di significatività osservata per confronti multipli.

Test visualizzati. I confronti pairwise sono disponibili per i test LSD, Sidak, Bonferroni, Games-Howell, Tamhane (T2) e (T3), *C* di Dunnett e T3 di Dunnett. Per i test S-N-K, *b* di Tukey, Duncan, R-E-G-W *F*, R-E-G-W *Q* e Waller sono disponibili sottoinsiemi omogenei per test di intervallo. Il test Differenza significativa di Tukey, GT2 di Hochberg, il test di Gabriel e il test di Scheffé sono sia test di confronto multiplo che test di intervalli.

GLM – Univariato: Opzioni

In questa finestra di dialogo sono disponibili statistiche opzionali. Le statistiche vengono calcolate tramite un modello di effetti fissi.

Visualizzazione. Selezionare **Statistica descrittiva** per produrre medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle. La funzione **Stima della dimensione degli effetti** fornisce un valore eta al quadrato parziale per ciascun effetto e per ciascuna stima del parametro. La statistica eta al quadrato consente di ottenere la proporzione della variabilità totale attribuibile a un fattore. Selezionare **Potenza osservata** per ottenere la potenza del test nel caso in cui l'ipotesi alternativa sia basata sul valore osservato. Selezionare **Stime del parametro** per ottenere stime del parametro, errori standard, test *t*, intervalli di confidenza e la potenza osservata per ciascun test. Selezionare **Matrice dei coefficienti di contrasto** per ottenere la matrice **L**.

La funzione **Test di omogeneità** produce i test di Levene per l'omogeneità della varianza per ogni variabile dipendente su tutte le combinazioni di livello dei fattori tra soggetti, solo per i fattori tra soggetti. Le opzioni Grafici di diffusione vs. densità e Grafici dei residui risultano utili per la verifica di ipotesi sui dati. Se non è disponibile alcun fattore, questa opzione risulta disattivata. Selezionare **Grafici dei residui** per ottenere un grafico dei residui osservati, attesi e standardizzati per ciascuna variabile dipendente. Questi grafici risultano utili per l'analisi dell'ipotesi di uguaglianza della varianza. Selezionare **Mancanza di adattamento** per controllare se la relazione tra la variabile dipendente e le variabili indipendenti può essere descritta in modo adeguato dal modello. **Funzione Generale stimabile (s)** consente di costruire test di ipotesi personalizzati in base alle funzioni generali stimabili. Le righe in qualsiasi matrice di coefficiente di contrasto sono combinazioni lineari della funzione generale stimabile (s).

I **Test dell'eteroschedasticità** sono disponibili per verificare se la varianza di errore (per ogni variabile dipendente) dipende dai valori delle variabili indipendenti. Per il **Test di Breusch-Pagan**, **Test di Breusch-Pagan modificato** e **Test F** è possibile specificare il modello su cui si basa il test. Per impostazione predefinita, il modello consiste in un termine costante, un termine che è lineare nei valori previsti, un termine che è quadratico nei valori previsti e un termine di errore.

Stime parametro con errori standard robusti visualizza una tabella di stime parametro insieme a errori standard robusti o di congruenza con l'eteroschedasticità, e le statistiche *t*, i valori di significatività e gli intervalli di confidenza utilizzati per gli errori standard robusti. Sono disponibili cinque diversi metodi per la stima della matrice di covarianza robusta.

HC0
: In base al campione ampio o asintotico originale, lo stimatore robusto, empirico o "sandwich" della matrice di covarianza delle stime parametro. La parte intermedia del sandwich contiene residui di minimi quadrati ordinari o pesati.

HC1
: Una modifica di campione finito HC0, moltiplicato per N/(N-p), dove N è la dimensione del campione e p è il numero dei parametri non ridondanti nel modello.

HC2
: Una modifica HC0 che coinvolge la divisione del residuo quadrato per 1-h, dove h è la leva per il caso.

HC3
: Una modifica HC0 che si approssima a uno stimatore jackknife. I residui quadrati sono divisi per il quadrato di 1-h.

HC4
: Una modifica HC0 che divide i residui quadrati per 1-h a una potenza che varia in base a h, N e p, con un limite superiore di 4.

Livello di significatività. Potrebbe risultare utile adattare il livello di significatività usato nei test post hoc e il livello di confidenza usato per la costruzione degli intervalli di confidenza. Il valore specificato viene inoltre usato per il calcolo della potenza osservata per il test. Quando si specifica un livello di significatività, nella finestra di dialogo viene visualizzato il livello di intervalli di confidenza associato.

Funzioni aggiuntive del comando UNIANOVA

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare gli effetti nidificati nel disegno (tramite il sottocomando `DESIGN`).
- Specificare test di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando `TEST`).
- Specificare contrasti multipli (tramite il sottocomando `CONTRAST`).
- Includere valori mancanti definiti dall'utente (tramite il sottocomando `MISSING`).
- Specificare criteri EPS (tramite il sottocomando `CRITERIA`).
- Creare una matrice **L**, una matrice **M** o una matrice **K** personalizzata (utilizzando i sottocomandi `LMATRIX`, `MMATRIX` e `KMATRIX`).
- Per i contrasti deviazione e i contrasti semplici, specificare una categoria di riferimento intermedia (tramite il sottocomando `CONTRAST`).

- Specificare metrica per contrasti polinomiali (tramite il sottocomando CONTRAST).
- Specificare termini di errore per confronti post-hoc (tramite il sottocomando POSTHOC).
- Calcolare medie marginali stimate per qualsiasi fattore o interazione tra fattori nell'elenco dei fattori (tramite il sottocomando EMMEANS).
- Assegnare un nome alle variabili temporanee (tramite il sottocomando SAVE).
- Costruire un file di dati matrice di correlazione (tramite il sottocomando OUTFILE).
- Costruire un file di dati matrice contenente statistiche derivate dai dati della tabella ANOVA tra soggetti (tramite il sottocomando OUTFILE).
- Salvare la matrice del disegno in un nuovo file di dati (tramite il sottocomando OUTFILE).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

GLM – Univariato: Salva

È possibile salvare i valori attesi dal modello, le misure correlate e i residui come nuove variabili nell'Editor dei dati. Molte di queste variabili possono essere usate per l'esame di ipotesi sui dati. Per salvare i valori in modo da poterli usare in un'altra sessione IBM SPSS Statistica , è necessario salvare il file di dati corrente.

Valori previsti. Valori attesi dal modello per ciascun caso.

- *Non standardizzati*.Il valore che il modello prevede per la variabile dipendente.
- *Pesato*.Valori attesi ponderati non standardizzati. Disponibile solo se era stata precedentemente selezionata una variabile WLS.
- *Errore standard*.Una stima della deviazione standard del valore medio della variabile dipendente per i casi che hanno gli stessi valori delle variabili indipendenti.

Diagnostiche. Misure per l'identificazione dei casi con combinazioni di valori insolite per le variabili indipendenti e dei casi che possono avere una notevole influenza sul modello.

- *Distanza di Cook*.Una misura di quanto cambierebbero i residui di tutti i casi se un particolare caso fosse escluso dal calcolo dei coefficienti di regressione. Un valore alto del D di Cook indica che l'esclusione di un caso dal calcolo delle statistiche di regressione modifica sostanzialmente i coefficienti.
- *Valori di leva*.Valori di leva non centrati. L'influenza relativa di ogni osservazione sull'adattamento del modello.

Residui. Un residuo non standardizzato corrisponde al valore effettivo della variabile dipendente diminuito del valore atteso dal modello. Sono inoltre disponibili residui standardizzati, studentizzati ed eliminati. Se è stata selezionata una variabile WLS, saranno inoltre disponibili residui non standardizzati pesati.

- *Non standardizzati*.La differenza tra un valore osservato e il valore stimato dal modello.
- *Pesato*.Residui ponderati non standardizzati. Disponibile solo se era stata precedentemente selezionata una variabile WLS.
- *Standardizzati*.Il residuo diviso per una stima della sua deviazione standard. I residui standardizzati, noti anche come residui di Pearson, hanno una media pari a 0 e una deviazione standard pari a 1.
- *Studentizzati*.Il residuo diviso per una stima della sua deviazione standard che varia da caso a caso, a seconda della distanza dei valori di ogni caso sulle variabili indipendenti dai mezzi delle variabili indipendenti. A volte indicato come residui internamente studentizzati.
- *Eliminato*.Il residuo per un caso quando tale caso è escluso dal calcolo dei coefficienti di regressione. È la differenza tra il valore della variabile dipendente e il valore previsto adattato.

Statistiche dei coefficienti. Scrive una matrice della varianza-covarianza delle stime del parametro nel modello in un nuovo dataset della sessione attiva o in un file di dati IBM SPSS Statistica esterno. Per ciascuna variabile dipendente è inoltre disponibile una riga di stime del parametro, una riga di errori standard delle stime del parametro, una riga di valori di significatività per le statistiche *t* corrispondenti alle stime del parametro e una riga di gradi di libertà dei residui. Per ciascuna variabile dipendente

di modelli multivariati sono disponibili righe simili. Quando è selezionato Statistiche congruenti con l'eteroschedasticità (disponibile solo per modelli univariata), la matrice varianza-covarianza è calcolata utilizzando uno stimatore robusto, la riga di errori standard visualizza gli errori standard robusti e i valori di significatività riflettono gli errori robusti. Il file matrice può essere usato in altre procedure che leggono i file matrice.

Medie marginali stimate GLM

Selezionare i fattori e le interazioni per cui si desiderano le stime delle medie marginali della popolazione nelle celle. Queste medie vengono adattate per le eventuali covariate.

Confronta effetti principali
Consente di eseguire confronti pairwise senza correzione tra le medie marginali stimate di qualsiasi effetto principale del modello, per fattori sia tra soggetti che entro soggetti. Questa opzione è disponibile solo se nell'elenco Medie marginali per sono stati selezionati effetti principali.

Confronta effetti principali semplici
L'impostazione è abilitata ogni qualvolta l'elenco di destinazione contiene uno o più effetti di prodotto o interazione (ad esempio A*B, A*B*C). L'impostazione supporta la specifica dei confronti tra semplici effetti principali, che sono effetti principali nidificati all'interno dei livelli di altri fattori.

Adattamento intervallo di confidenza
Selezionare la differenza meno significativa (LSD), l'adattamento di Bonferroni o di Sidak agli intervalli di confidenza e la significatività. Questo articolo è disponibile solo quando viene selezionato **Confronta gli effetti principali** e / o **Confronta semplici effetti principali** .

Specifica delle medie marginali stimate

1. Dai menu scegliere una delle procedure disponibili sotto > **Analizza** > **Modello lineare generale**.
2. Nella finestra d dialogo principale, fare clic su **Medie EM**.

GLM – Univariato: Opzioni

In questa finestra di dialogo sono disponibili statistiche opzionali. Le statistiche vengono calcolate tramite un modello di effetti fissi.

Visualizzazione. Selezionare **Statistica descrittiva** per produrre medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle. La funzione **Stima della dimensione degli effetti** fornisce un valore eta al quadrato parziale per ciascun effetto e per ciascuna stima del parametro. La statistica eta al quadrato consente di ottenere la proporzione della variabilità totale attribuibile a un fattore. Selezionare **Potenza osservata** per ottenere la potenza del test nel caso in cui l'ipotesi alternativa sia basata sul valore osservato. Selezionare **Stime del parametro** per ottenere stime del parametro, errori standard, test *t*, intervalli di confidenza e la potenza osservata per ciascun test. Selezionare **Matrice dei coefficienti di contrasto** per ottenere la matrice **L**.

La funzione **Test di omogeneità** produce i test di Levene per l'omogeneità della varianza per ogni variabile dipendente su tutte le combinazioni di livello dei fattori tra soggetti, solo per i fattori tra soggetti. Le opzioni Grafici di diffusione vs. densità e Grafici dei residui risultano utili per la verifica di ipotesi sui dati. Se non è disponibile alcun fattore, questa opzione risulta disattivata. Selezionare **Grafici dei residui** per ottenere un grafico dei residui osservati, attesi e standardizzati per ciascuna variabile dipendente. Questi grafici risultano utili per l'analisi dell'ipotesi di uguaglianza della varianza. Selezionare **Mancanza di adattamento** per controllare se la relazione tra la variabile dipendente e le variabili indipendenti può essere descritta in modo adeguato dal modello. **Funzione Generale stimabile (s)** consente di costruire test di ipotesi personalizzati in base alle funzioni generali stimabili. Le righe in qualsiasi matrice di coefficiente di contrasto sono combinazioni lineari della funzione generale stimabile (s).

I **Test dell'eteroschedasticità** sono disponibili per verificare se la varianza di errore (per ogni variabile dipendente) dipende dai valori delle variabili indipendenti. Per il **Test di Breusch-Pagan**, **Test di Breusch-Pagan modificato** e **Test F** è possibile specificare il modello su cui si basa il test. Per impostazione predefinita, il modello consiste in un termine costante, un termine che è lineare nei valori previsti, un termine che è quadratico nei valori previsti e un termine di errore.

Stime parametro con errori standard robusti visualizza una tabella di stime parametro insieme a errori standard robusti o di congruenza con l'eteroschedasticità, e le statistiche *t*, i valori di significatività e gli intervalli di confidenza utilizzati per gli errori standard robusti. Sono disponibili cinque diversi metodi per la stima della matrice di covarianza robusta.

HC0
In base al campione ampio o asintotico originale, lo stimatore robusto, empirico o "sandwich" della matrice di covarianza delle stime parametro. La parte intermedia del sandwich contiene residui di minimi quadrati ordinari o pesati.

HC1
Una modifica di campione finito HC0, moltiplicato per N/(N-p), dove N è la dimensione del campione e p è il numero dei parametri non ridondanti nel modello.

HC2
Una modifica HC0 che coinvolge la divisione del residuo quadrato per 1-h, dove h è la leva per il caso.

HC3
Una modifica HC0 che si approssima a uno stimatore jackknife. I residui quadrati sono divisi per il quadrato di 1-h.

HC4
Una modifica HC0 che divide i residui quadrati per 1-h a una potenza che varia in base a h, N e p, con un limite superiore di 4.

Livello di significatività. Potrebbe risultare utile adattare il livello di significatività usato nei test post hoc e il livello di confidenza usato per la costruzione degli intervalli di confidenza. Il valore specificato viene inoltre usato per il calcolo della potenza osservata per il test. Quando si specifica un livello di significatività, nella finestra di dialogo viene visualizzato il livello di intervalli di confidenza associato.

Modello di regressione ausiliaria GLM

La finestra di dialogo Modello di regressione ausiliaria specifica il modello utilizzato per verificare l'eteroschedasticità.

Utilizza valori previsti
Utilizza un modello che consiste in un termine costante, un termine che è lineare nei valori previsti, un termine che è quadratico nei valori previsti e un termine di errore.

Utilizza modello univariata
Utilizza il modello specificato nella finestra secondaria Modello. Un termine intercetta è incluso se il modello spcificato non ne contiene uno.

Modello personalizzato
Utilizza il modello specificato esplicitamente dall'utente.

Crea termini
Utilizzare questa opzione quando si desidera includere termini non nidificati di un certo tipo (ad esempio, effetti principali) per tutte le combinazioni di una serie selezionata di fattori e covariate.

Crea termini personalizzati
Utilizzare questa opzione quando si desidera includere termini nidificati o creare esplicitamente una variabile termine dalla variabile. La creazione di un termine nidificato richiede le seguenti operazioni:

Funzioni aggiuntive del comando UNIANOVA

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare gli effetti nidificati nel disegno (tramite il sottocomando `DESIGN`).
- Specificare test di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando `TEST`).
- Specificare contrasti multipli (tramite il sottocomando `CONTRAST`).
- Includere valori mancanti definiti dall'utente (tramite il sottocomando `MISSING`).

- Specificare criteri EPS (tramite il sottocomando CRITERIA).
- Creare una matrice **L**, una matrice **M** o una matrice **K** personalizzata (utilizzando i sottocomandi LMATRIX, MMATRIX e KMATRIX).
- Per i contrasti deviazione e i contrasti semplici, specificare una categoria di riferimento intermedia (tramite il sottocomando CONTRAST).
- Specificare metrica per contrasti polinomiali (tramite il sottocomando CONTRAST).
- Specificare termini di errore per confronti post-hoc (tramite il sottocomando POSTHOC).
- Calcolare medie marginali stimate per qualsiasi fattore o interazione tra fattori nell'elenco dei fattori (tramite il sottocomando EMMEANS).
- Assegnare un nome alle variabili temporanee (tramite il sottocomando SAVE).
- Costruire un file di dati matrice di correlazione (tramite il sottocomando OUTFILE).
- Costruire un file di dati matrice contenente statistiche derivate dai dati della tabella ANOVA tra soggetti (tramite il sottocomando OUTFILE).
- Salvare la matrice del disegno in un nuovo file di dati (tramite il sottocomando OUTFILE).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Bland-Altman

L'analisi di Bland Altman è una tecnica grafica per valutare il bias tra le differenze medie. Altman e Bland hanno proposto questa analisi nel 1983. L'analisi di Bland Altman si basa sulla quantificazione dell'accordo tra due misure quantitative esplorando le differenze medie e costruendo i limiti di accordo.

Esempio

I laboratori medici devono spesso valutare l'accordo tra due metodi di misurazione, scegliere tra due metodi o valutare un metodo nuovo o alternativo. Inoltre, quando si verifica un problema di allineamento tra due strumenti, sono necessari alcuni strumenti per misurare e valutare le differenze e la causa di tali differenze.

Questa procedura richiede Statistics Base Edition.

Dai menu scegliere: **Analisi** > **Statistiche descrittive** > **Analisi di Bland Altman**...

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

Considerazioni sui dati: Le variabili sono in cifre.

1. Selezionate le variabili e aggiungetele alla cartella **Variabili accoppiate:**.
2. Se si seleziona **OK**, la procedura viene eseguita considerando le impostazioni predefinite.
3. È possibile scegliere tra **Applica la trasformazione dei log** o **Solo in coppia** o entrambi.
4. Fare clic su **OK**.

Applicare la trasformazione log: si consiglia la trasformazione **logaritmica** se le differenze non sono distribuite normalmente.

Solo accoppiati: questa opzione viene specificata quando la prima variabile della prima colonna viene confrontata con la prima variabile della seconda colonna.

Nota: Tra le variabili accoppiate, l'analisi viene effettuata solo fino all'ultimo valore dell'insieme di variabili che contiene il minor numero di valori. L'analisi viene eseguita completamente per gli insiemi di variabili accoppiate contenenti lo stesso numero di valori.

Bland Altman: criteri

È possibile modificare le impostazioni predefinite nella scheda **Bland Atman: Criteri**.

1. Per impostazione predefinita, l'opzione di mostrare i rapporti in percentuale non è selezionata. Se si desidera aggiungerlo, selezionare **Rapporto in percentuale**.
2. Controllare le impostazioni di **Ambito di applicazione** e **Limiti dell'accordo** e modificarle se necessario.
3. Fare clic su **Continua**.

Bland Altman: Stampa

1. Selezionare le opzioni che si desidera includere nell'output.
2. Se si seleziona **Stampa dettagli dell'analisi**, l'impostazione predefinita per il numero massimo di righe da visualizzare è 10. Modificare il numero per ottenere un numero maggiore o minore di righe nella visualizzazione. Il numero massimo di righe di dati visualizzabili è 1000. **Ordinamento** può essere **basato sulle** quattro opzioni dell'elenco e sull'**ordine ascendente o discendente**.

   Per impostazione predefinita, **Ordinamento** è impostato in base alla prima misura e in ordine decrescente.
3. Fare clic su **Continua**.

Bland Altman: trama

È possibile modificare le impostazioni predefinite nella scheda **Analisi dell'Atman Bland: Traccia**.

1. Scegliere le impostazioni desiderate, tra cui **Piani di rapporto**, la **linea di riferimento a differenza zero**, **limiti di accordo** e **intervalli di confidenza delle differenze medie**.
2. Fare clic su **Continua**.

Test del chi-quadrato

La procedura del test chi-quadro tabula una variabile in categorie e calcola una statistica chi-quadro. Il test del chi-quadro è un test statistico di ipotesi utilizzato per determinare se esiste un'associazione significativa tra due variabili categoriali. Il test valuta se le frequenze osservate nei dati differiscono dalle frequenze attese in base all'ipotesi nulla, che presuppone l'assenza di associazione tra le variabili.

Questa procedura richiede Statistics Base Edition.

Esempio
Nelle ricerche di mercato, gli analisti devono comprendere le preferenze dei consumatori. A tal fine, confrontano il comportamento dei clienti osservato con i modelli previsti per creare strategie di marketing efficaci.

Statistiche
Il chi-quadro di Pearson, il chi-quadro di verosimiglianza, il test esatto di Fisher e il chi-quadro corretto di Yates (correzione di continuità).

Considerazioni e ipotesi sui dati
Utilizzare variabili categoriali numeriche ordinate o non ordinate (livelli di misurazione ordinali o nominali). Per convertire le variabili stringa in variabili numeriche, utilizzare il comando Ricodifica automatica del menu Trasforma.

I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Si presume che i dati rappresentino un campione casuale. Le frequenze attese per ogni categoria devono essere almeno pari a 1. Non più del 20% delle categorie dovrebbe avere frequenze previste inferiori a 5.

Dal menu, fare clic su **Analizza** > **statistiche** > **descrittiveChiSquare**.

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

1. Selezionare le variabili e aggiungerle alle **righe** e alle **colonne**.
2. Fare clic su **OK**. La procedura viene eseguita considerando le impostazioni predefinite.

Test del chi-quadro: Esatta

- *Asintotico*. Il livello di significatività basato sulla distribuzione asintotica di una statistica di test. Generalmente, un valore inferiore a 0.05 è considerato significativo. La significatività asintotica si basa sull'assunto che il dataset è grande. Se il dataset è piccolo o distribuito male, questa potrebbe non essere una buona indicazione della significatività.
- *Stima Monte Carlo*. Una stima imparziale del livello di significatività esatto, calcolata campionando ripetutamente da un insieme di tabelle di riferimento con le stesse dimensioni e margini di riga e colonna della tabella osservata. Il metodo Monte Carlo consente di stimare la significatività esatta anche in assenza delle ipotesi necessarie per il metodo asintotico Questo metodo è particolarmente utile quando il dataset di cui si dispone è troppo ampio per calcolare la significatività esatta ma i dati non soddisfano i presupposti del metodo asintotico.
- *Esatto*. La probabilità del risultato osservato o uno più estremo viene calcolata in maniera esatta. Di norma, un livello di significatività inferiore a 0,05 è considerato significativo, indicando che c'è qualche relazione tra le variabili riga e colonna.

Test del chi-quadro: Celle

La procedura del chi-quadro visualizza le frequenze previste. Ogni cella della tabella può contenere qualsiasi combinazione di conteggi e percentuali.

Conteggi
Il numero di casi effettivamente osservati e il numero di casi attesi se le variabili di riga e colonna sono indipendenti tra loro. È possibile scegliere di nascondere i conteggi che sono inferiori a un numero intero specificato. I valori nascosti vengono visualizzati come **<N**, dove **N** è il numero intero specificato. Il numero intero specificato deve essere superiore o uguale a 2, ma è consentito utilizzare il valore 0 per specificare che non viene nascosto alcun conteggio.

Percentuali
Le percentuali possono essere aggiunte nelle righe o nelle colonne. Sono disponibili anche le percentuali del numero totale di casi rappresentati nella tabella (un solo strato).

Nota: Se si seleziona **Nascondi conteggi piccoli** nel gruppo Conteggi, vengono nascoste anche le percentuali associate ai conteggi nascosti.

Crea tabella di stile APA
Crea tabelle di output conformi alle linee guida dello stile APA.

Nota: Le opzioni **Osservato**, **Previsto**, **Riga**, **Colonna** e **Totale** non sono disponibili quando si seleziona **Crea tabella stile APA**.

Test del chi-quadro: Formato

Con l'opzione Formato è possibile specificare l'ordine delle variabili visualizzate nei risultati del test.

Vai a **Analizza** > **statistiche descrittive** > **Chiquadrato**. Fare clic su **Formato** nella finestra di dialogo del **test del chi-quadro**.

È possibile disporre le righe nell'ordine crescente o decrescente dei valori della variabile di riga. In **Ordine di riga**, selezionare **Ascendente** o **Discendente**.

Correlazioni di bivariate

La procedura Correlazioni bivariate consente di calcolare coefficiente di correlazione di Pearson, rho di Spearman e tau-*b* di Kendall con i rispettivi livelli di significatività. Le correlazioni consentono di misurare la relazione tra variabili o punteggi di rango. Prima di calcolare un coefficiente di correlazione, è necessario valutare la presenza di valori anomali nei dati (che possono causare risultati errati) e l'esistenza di una relazione lineare. Il coefficiente di correlazione di Pearson è una misura di associazione lineare. Due variabili possono essere perfettamente correlate, ma se la relazione non è lineare il coefficiente di correlazione di Pearson non è la statistica migliore per misurare tale associazione.

Le impostazioni dell'intervallo di confidenza sono disponibili per Pearson e Spearman.

Esempio
Il numero di partite vinte da una squadra di baseball è correlato con la media dei punti totalizzati per ciascuna partita? Un grafico a dispersione indica l'esistenza di una relazione lineare. Analizzando i dati della stagione NBA 1994–1995 si deduce che il coefficiente di correlazione di Pearson (0.581) è significativo a livello 0.01. Si può presumere che il numero di partite vinte per stagione sia inversamente proporzionale ai punti totalizzati dagli avversari. Queste variabili sono correlate in modo negativo (–0.401) e la correlazione è significativa a livello 0.05.

Statistiche
Per ogni variabile: numero di casi con valori non mancanti, media e deviazione standard. Per ogni coppia di variabili: il coefficiente di correlazione di Pearson, rho di Spearman, tau-*b* di Kendall, cross-product delle deviazioni e covarianza.

Considerazioni sui dati

Dati
Utilizzare le variabili quantitative simmetriche per il coefficiente di correlazione di Pearson e le variabili quantitative o le variabili con categorie ordinate per rho di Spearman e tau-*b* di Kendall.

Ipotesi
Il coefficiente di correlazione di Pearson assume che ciascuna coppia di variabili sia bivariata normale.

Ottenere Correlazioni Bivariate

Dai menu, scegliere:

Analizza > **Correlazione** > **Bivariata...**

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

1. Selezionare una o più variabili numeriche.

   Sono inoltre disponibili le seguenti opzioni:

   Coefficienti di correlazione
   Per le variabili quantitative, normalmente distribuite, scegliere il coefficiente di correlazione **Pearson** . Se i dati non sono normalmente distribuiti o prevedono categorie ordinate, scegliere **Tau-b di Kendall** o **Spearman**, per misurare l'associazione tra punteggi di rango. Il valore dei coefficienti di correlazione è compreso tra –1 (una relazione negativa perfetta) e +1 (una relazione positiva perfetta). Il valore 0 indica l'assenza di relazione lineare. Interpretando i risultati, evitare di trarre conclusioni di tipo causa-effetto sulla base di una correlazione significativa.

   Test di significatività
   È possibile selezionare le probabilità a due code o a una tazzina. Se si conosce in anticipo la direzione dell'associazione, selezionare **A una coda.**. In alternativa, selezionare **A due code**.

   Evidenzia correlazioni significative
   I coefficienti di correlazione significativi al livello 0,05 vengono identificati con un asterisco singolo e quelli significativi al livello 0,01 con due asterischi.

   Mostra solo il triangolo inferiore
   Quando viene selezionato, nell'output viene presentato solo il triangolo inferiore della tabella di correlazione. Quando non viene selezionato, la tabella di matrice di correlazione completa viene presentata nell'output. L'impostazione consente l'output della tabella per aderire alle linee guida in stile APA.

   Mostra diagonale
   Quando viene selezionato, nell'output viene presentato il triangolo inferiore del tavolo della matrice di correlazione insieme ai valori diagonali. L'impostazione consente l'output della tabella per aderire alle linee guida in stile APA.
2. È possibile opzionalmente selezionare quanto segue:

- Fare clic su **Opzioni ...** per specificare le statistiche di correlazione di Pearson e le impostazioni dei valori mancanti.
- Clicca su **Stile ...** per specificare le condizioni per la modifica automatica delle proprietà delle tabelle pivot in base a specifiche condizioni.
- Fare clic su **Bootstrap ...** per derivare stime robusta di errori standard e intervalli di confidenza per stime quali la media, la mediana, la proporzione, il rapporto di probabilità, il coefficiente di correlazione o il coefficiente di regressione.
- Clicca **Intervallo di confidenza ...** per impostare le opzioni per la stima degli intervalli di confidenza.

Correlazioni bivariate: Opzioni

Statistiche

Per le correlazioni di Pearson è possibile scegliere uno o entrambi i seguenti:

Medie e deviazioni standard

Visualizzate per ciascuna variabile. Viene indicato anche il numero di casi con valori non mancanti. I valori mancanti vengono gestiti variabile per variabile indipendentemente dall'impostazione corrispondente.

Deviazioni e covarianze cross-product

Viene visualizzato per ciascuna coppia di variabili. Il cross-product delle deviazioni è equivalente alla somma dei prodotti delle variabili corrette per la media. È il numeratore del coefficiente di correlazione di Pearson. La covarianza è una misura non standardizzata della relazione tra due variabili, uguale alla deviazione cross-product divisa per *N*–1.

Valori mancanti

È possibile scegliere tra le opzioni seguenti:

Escludi casi pairwise

I casi con valori mancanti per una o entrambe le variabili di una coppia per un coefficiente di correlazione vengono esclusi dall'analisi. Poiché ciascun coefficiente si basa su tutti i casi con codici validi per quella particolare coppia di variabili, in tutti i calcoli verrà utilizzato il maggior numero di informazioni disponibile. In questo modo è possibile ottenere un insieme di coefficienti basati su un numero variabile di casi.

Escludi casi listwise

I casi con valori mancanti per qualsiasi variabile vengono esclusi da tutte le correlazioni.

Intervallo Di Confidenza Correlazioni Bivariate

Dialogo Intervallo di confidenza fornisce opzioni per la stima degli intervalli di confidenza. La finestra di dialogo è disponibile quando **Pearson**, **Kendall's tau - b**o **Spearman** è selezionato sulla finestra di dialogo Correlazioni Bivariate.

Stima dell'intervallo di confidenza del parametro di correlazione bivariata

Controlla la stima dell'intervallo di confidenza del parametro di correlazione bivariata. Quando viene selezionata, si verifica l'intervallo di confidenza.

Intervallo di confidenza (%)

Specifica il livello di confidenza per tutti gli intervalli di confidenza prodotti. Specificare un valore numerico compreso tra 0 e 100. 95 è il valore predefinito.

Correlazione di Pearson

L'impostazione **Applica l'adattamento bias** controlla se viene applicata la regolazione del bias. Per impostazione predefinita, l'impostazione non è selezionata, che non prende in considerazione il termine bias. Quando viene selezionata, viene applicata la regolazione del bias alla stima dei limiti di confidenza. L'impostazione è disponibile quando **Pearson** viene selezionato sulla finestra di dialogo Correlazioni Bivariate.

Correlazione di Spearman

L'impostazione è disponibile quando **Spearman** viene selezionato sulla finestra di dialogo Correlazioni Bivariate e fornisce opzioni per stimare la varianza Spearman Correlazione attraverso i seguenti metodi:

- **Fieller, Hartley e Pearson**
- **Bonett e Wright**
- **Coruso e Cliff**

Funzioni aggiuntive dei comandi CORRELATIONS e NONPAR CORR

Il linguaggio della sintassi dei comandi consente inoltre di:

- Scrivere una matrice di correlazione per la correlazione di Pearson da utilizzare in luogo dei dati grezzi per ottenere altri tipi di analisi, ad esempio l'analisi fattoriale (con il sottocomando `MATRIX`).
- Ottenere correlazioni di ciascuna variabile presente in un elenco con le variabili corrispondenti presenti in un secondo elenco (utilizzando la parola chiave `WITH` con il sottocomando `VARIABLES`).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Associazione di prossimità

Questa documentazione guida gli utenti nella specificazione dei dati di input, nella selezione del modello, nelle funzioni di trasformazione, nelle impostazioni di dimensionalità e nell'inclusione di attributi e proprietà nella Mappatura di prossimità.

La mappatura di prossimità è una tecnica di visualizzazione utilizzata per ridurre la dimensionalità dei dati multivariati e per visualizzare le relazioni tra gli oggetti (casi, elementi o altre entità) in una configurazione spaziale. Sviluppato originariamente con il nome di multidimensional scaling (MDS) in psicologia, il metodo utilizza le prossimità (somiglianze, dissomiglianze o distanze) per posizionare gli oggetti in modo che la loro disposizione spaziale rifletta la struttura dei dati. Gli oggetti simili appaiono vicini tra loro, mentre quelli dissimili sono distanziati. La mappatura di prossimità supporta l'interpretazione e il riconoscimento dei modelli trasformando dati relazionali complessi in grafici intuitivi a due o tre dimensioni.

In IBM SPSS Statistica , la mappatura di prossimità è implementata attraverso la procedura PROXMAP. A differenza di PROXSCAL, che si limita ad analizzare le matrici di prossimità, PROXMAP accetta diverse forme di input e offre una maggiore flessibilità. Supporta più fonti di prossimità, incorpora variabili aggiuntive (attributi e proprietà) e offre un'ampia gamma di opzioni di trasformazione e restrizione. Queste caratteristiche rendono PROXMAP uno strumento potente per l'esplorazione della struttura multivariata e per l'integrazione di diversi tipi di dati in una rappresentazione spaziale unificata.

Quando le informazioni di prossimità sono disponibili da più fonti, specificate direttamente o derivate da dati multivariati, PROXMAP può adattare diversi modelli di differenze individuali oltre al modello di *identità* predefinito. I modelli disponibili includono il modello dell'identità ponderata (dilatazione), il modello euclideo ponderato (diagonale) e il modello euclideo generalizzato con restrizioni di rango opzionali.

Rispetto alle tecniche tradizionali di riduzione delle dimensioni, come l'analisi delle componenti principali (PCA) e l'MDS classico, PROXMAP fornisce una rappresentazione più diretta e flessibile delle prossimità. L'MDS classico può essere usato per inizializzare la configurazione. Se applicato a dati multivariati, l'MDS classico è equivalente all'analisi delle coordinate principali, una variante della PCA.

PROXMAP accetta i seguenti tipi di input.

Tipo di input	Descrizione
Prossimità grezze e derivate	Le prossimità derivate sono calcolate da variabili numeriche, ordinali o nominali. Entrambi i tipi di

Tipo di input	Descrizione
	prossimità vengono facoltativamente trasformati utilizzando funzioni monotoniche (ad esempio, ordinali o basate su spline), prima di essere approssimati dalle distanze.
Attributi	Variabili che vincolano la configurazione e che appaiono come direzioni nello spazio dei giunti (mappatura supervisionata).
Proprietà	Variabili supplementari utilizzate per l'interpretazione; sono visualizzate nello spazio ma non influenzano la configurazione (mappatura non supervisionata).

Una caratteristica unica di PROXMAP è la possibilità di utilizzare lo stesso set di variabili in ruoli diversi:

- Come base per la derivazione di prossimità
- Come attributi che modellano la configurazione
- Come proprietà che supportano l'interpretazione

In alternativa, si possono assegnare set di variabili distinti a ciascun ruolo.

Oltre alla trasformazione lineare, PROXMAP offre varie opzioni di trasformazione (tra cui trasformazioni ordinali, spline monotoniche, spline non monotoniche e nominali) sia per gli attributi che per le proprietà. Per le prossimità, le trasformazioni devono essere monotone e comprendono anche le trasformazioni di potenza. L'output include diagnostici grafici come diagrammi di stress, alberi di minima estensione e grafici di prossimità per facilitare l'interpretazione e la valutazione.

Esempio

La mappatura di prossimità può essere ampiamente applicata in diverse discipline, tra cui la psicologia e le scienze comportamentali (percezione, memoria, preferenze), il marketing (mappatura dei marchi, segmentazione), la sociologia (mappatura delle ideologie, reti sociali), l'istruzione e la psicometria (diagnostica dei test, profilazione delle competenze), la linguistica (somiglianza semantica, mappatura dei dialetti), Geografia (distanze percepite, analisi spaziale ecologica), Genomica e Bioinformatica (modelli di espressione, struttura della popolazione, esplorazione dell'eterogeneità dell'espressione di una singola cellula), Chemioinformatica (somiglianza molecolare), Informatica (visualizzazione di embedding), Antropologia e Archeologia (somiglianza di artefatti, dati cranici).

Dati

I dati da analizzare comprendono variabili che rappresentano una matrice (o matrici) di prossimità, oppure rappresentano dati multivariati che vengono convertiti in matrice o matrici di prossimità.

Per i dati multivariati, le variabili possono essere un misto di variabili di scala, ordinali e nominali. Il numero minimo di variabili necessarie è 3

Per i dati di prossimità, fornire un numero di variabili pari al numero di casi.

Sia per i dati di prossimità che per i dati multivariati, il numero minimo di casi necessari è 3.

L'output include configurazioni spaziali, vari grafici diagnostici e statistiche di adattamento. I biplot visualizzano gli oggetti insieme a variabili, attributi e proprietà in uno spazio comune.

Per avviare l'analisi PROXMAP, fare clic su **Analizza** > **Mappatura** > **Proximity Mapping**.

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

Mappatura di prossimità: Dati

Nella scheda **Dati** è possibile definire una o più fonti di prossimità basate sulle variabili del set di dati. Ogni fonte viene creata selezionando le variabili dal set di dati attivo e specificando se rappresentano matrici di prossimità o dati multivariati da cui derivare le prossimità.

Variabili
L'elenco **Variabili** visualizza tutte le variabili del dataset attivo. Le variabili comprendono variabili numeriche, ordinali e nominali che possono essere utilizzate nel calcolo della prossimità. Selezionare una o più variabili e spostarle in **Variabili nell'origine**.

Variabili nella sorgente
Le **Variabili nella sorgente** elencano le variabili attualmente selezionate per la sorgente di prossimità attiva. Queste variabili definiscono la struttura della sorgente e vengono utilizzate per ricavare le prossimità (per i dati multivariati) o per specificare una matrice di prossimità (se le prossimità sono fornite direttamente). Utilizzare i pulsanti freccia su o giù per riordinare le variabili o rimuoverle se necessario.

Etichette oggetto
Con le **Etichette oggetto** è possibile specificare una variabile che fornisce le etichette per gli oggetti (casi) da visualizzare nei grafici e nelle tabelle di output. Selezionate una variabile dall'elenco e spostatela in **Etichette oggetto** per assegnarla.

definizione di origine
A destra delle **Variabili in Origine**, i campi Nome dell'origine e Etichetta dell'origine definiscono l'identificatore interno e l'etichetta descrittiva della prossimità corrente source.The Il **nome dell'origine** deve essere un identificatore unico preceduto da *$*.

Per gestire le sorgenti, utilizzare i seguenti pulsanti.

- **Aggiungi origine** crea una nuova origine di prossimità dalle variabili presenti nel riquadro Variabili nell'origine.
- **Cambia sorgente** modifica la sorgente attualmente selezionata utilizzando la selezione aggiornata della variabile.
- **Rimuovi sorgente** elimina dall'elenco la sorgente attualmente selezionata.

Tipo di dati
Utilizzare il selettore **Tipo di dati** per specificare l'interpretazione delle variabili selezionate. Con il gruppo **Tipo di dati** è possibile indicare il modo in cui le prossimità sono derivate dalle variabili selezionate:

- **Dati di prossimità** Utilizzare questa opzione se l'input è già una matrice di prossimità.
- **Dati multivariati** (impostazione predefinita) tratta le variabili selezionate come caratteristiche degli oggetti. PROXMAP ricava le prossimità calcolando le distanze tra gli oggetti in base a queste variabili.
  - **Euclidea** : Le prossimità si basano su distanze euclidee non standardizzate.
  - **Euclidea standardizzata** : Le variabili vengono standardizzate prima di calcolare le distanze (impostazione predefinita).

Pesi
Quando il **tipo di dati** è impostato su **Dati di prossimità**, è disponibile un elenco opzionale di **pesi di prossimità**. Qui è possibile specificare ulteriori variabili di peso da associare ai valori di prossimità. Questi pesi regolano l'influenza di ciascuna prossimità nel processo di ottimizzazione.

Mappatura di prossimità: Pesi

Nella scheda **Pesi** è possibile specificare i pesi di prossimità (cella) per ogni origine e i pesi degli oggetti (caso) per i dati di prossimità. Questi pesi influenzano il contributo delle prossimità al processo di ottimizzazione. Questa scheda è rilevante solo quando nella scheda **Dati** è selezionato il **tipo di dati Dati di prossimità**.

Variabili
L'elenco **Variabili** visualizza tutte le variabili del dataset attivo. Queste variabili devono contenere i pesi di prossimità. Selezionare una o più variabili e spostarle nella sezione **Pesi in Sorgente**. Queste variabili devono contenere i pesi di prossimità per una specifica sorgente.

Pesi in Fonte
Pesi nella sorgente visualizza le variabili di peso selezionate associate alla sorgente attualmente attiva. Queste variabili definiscono una matrice di pesi di prossimità che corrisponde alla struttura della matrice di prossimità corrispondente. Il numero di variabili di peso deve essere uguale al numero di variabili della matrice di origine.

Usare **Aggiungi pesi**, **Modifica pesi** o **Rimuovi pesi** per gestire le definizioni dei pesi per la sorgente selezionata in **Sorgenti**.

Origine/i
L'elenco **Sorgenti** mostra tutte le sorgenti di prossimità definite a cui è possibile assegnare pesi. Selezionare una sorgente per visualizzare o modificare i pesi associati.

Pesi oggetto (caso)
Con il campo Etichette oggetto sotto i riquadri di selezione, è possibile definire una variabile che contiene i pesi per i singoli casi (oggetti). Questi pesi regolano l'influenza di ciascun oggetto nella funzione di perdita quando si utilizzano i dati di prossimità. La ponderazione dei casi non è applicabile ai dati multivariati.

Funzione pesi

La **funzione Pesi** offre tre opzioni per speciali schemi di ponderazione di prossimità che sono funzioni dei valori di prossimità:

- **Nessuna** : Non viene applicata alcuna ponderazione speciale. Vengono utilizzati solo i pesi di prossimità e di oggetto specificati dall'utente (se presenti).
- **Sammon** : applica pesi alla vicinanza tra gli oggetti *i* e *j*. Questo peso enfatizza le piccole vicinanze, come nella mappatura di Sammon.
- **Energia** : Applica pesi che pongono ancora più enfasi sulle piccole vicinanze.

Il peso totale applicato a ogni prossimità è il prodotto dei pesi di prossimità definiti dall'utente, dei pesi dei casi e della funzione di peso speciale. Questi pesi derivati regolano l'importanza relativa di ogni prossimità nella funzione di perdita che viene minimizzata da PROXMAP.

Mappatura di prossimità: Iniziale

La scheda **Iniziale** specifica la configurazione iniziale del processo di ottimizzazione iterativa utilizzato in PROXMAP. Una buona configurazione iniziale può accelerare la convergenza e migliorare la qualità della soluzione. È possibile scegliere tra l'inizializzazione automatica utilizzando la scalatura classica sulle prossimità grezze o in base ai loro ordini di rango. È anche possibile fornire una configurazione personalizzata dal set di dati.

Scalabilità classica
Questa è l'opzione predefinita. Una soluzione scalare classica (Analisi delle coordinate principali) viene calcolata sulla base delle prossimità. Fornisce una configurazione di partenza derivata analiticamente nel numero di dimensioni specificato.

Due opzioni aggiuntive controllano il trattamento delle prossimità durante la scalatura classica:

- Con **Valore costante additivo**, è possibile specificare una costante fissa da aggiungere a tutte le prossimità prima di scalare. Questo può migliorare le proprietà metriche ed evitare autovalori negativi. Il valore non deve risultare in una prossimità negativa.
- Con l'opzione **Stima costante additiva**, PROXMAP stima automaticamente la costante ottimale da aggiungere. Ciò garantisce che la matrice di distanza risultante sia semidefinita positiva.
- Con **Use Rank Numbers for Proximities** è possibile applicare la scalatura classica agli ordini di rango delle prossimità piuttosto che ai valori grezzi. In questo modo si ottiene una configurazione di partenza `ordinal` e si riduce la sensibilità ai valori anomali.

Configurazione personalizzata

Con la **Configurazione personalizzata**, è possibile fornire direttamente le proprie coordinate iniziali. Le variabili del set di dati vengono selezionate come dimensioni della configurazione di partenza. Queste variabili devono contenere valori numerici validi senza dati mancanti.

L'elenco **delle variabili** contiene tutte le variabili disponibili. Selezionare le variabili che contengono la configurazione personalizzata e spostarle nelle **Dimensioni**.

L'elenco **Dimensioni** visualizza le variabili selezionate per essere utilizzate come coordinate iniziali. Il numero di variabili selezionate deve corrispondere al numero di dimensioni. Se vengono fornite più variabili, i valori vengono proiettati sulle prime *p* componenti principali, dove *p* è il numero di dimensioni.

Current Number of Dimensions visualizza la dimensionalità specificata per la soluzione finale (specificata nella scheda **Criteri**).

Mappatura di prossimità: Modello

Nella scheda **Modello** è possibile specificare il modello di differenze individuali da applicare quando si analizzano più fonti di prossimità. Definisce anche come vengono gestite le trasformazioni di prossimità tra le fonti: in modo uniforme (incondizionato) o indipendente (matrice-condizionato).

Modello a tre vie

Le impostazioni del **modello a tre vie** determinano il modo in cui le prossimità provenienti da più fonti vengono mappate in una rappresentazione spaziale comune, utilizzando i pesi spaziali. Il modello determina il modo in cui i pesi dello spazio vengono applicati allo spazio comune. Per alcuni modelli, i pesi spaziali sono equivalenti ai pesi spaziali. Quando viene definita una sola sorgente, tutte le opzioni tranne **Identità** sono disattivate.

Opzione	Descrizione
Identità	Presuppone che tutte le sorgenti condividano la stessa configurazione, senza scalature specifiche per la sorgente. Disponibile solo quando viene specificata una singola sorgente.
Dilatazione (identità ponderata)	Ogni sorgente è modellata con un unico fattore di scala applicato uniformemente a tutte le dimensioni. Con questa opzione, ogni fonte può percepire lo spazio comune a una scala complessiva diversa.
Euclidea ponderata (diagonale)	Ogni fonte assegna pesi diversi alle dimensioni dello spazio comune. Ciò consente alle sorgenti di enfatizzare le diverse caratteristiche spaziali.
Euclideo generalizzato	Ogni sorgente può avere una versione ruotata e ripesata (in scala) dello spazio comune. Questo è il modello più flessibile e si adatta a un'ampia gamma di differenze individuali. • **Rango completo** : Ogni fonte utilizza l'intera dimensionalità dello spazio comune. • **Rango ridotto** : Quando si applica un vincolo di rango ridotto, la configurazione di ciascuna sorgente (spazio individuale, conrango = numero di dimensioni - 1) è rappresentato in un sottospazio a bassa dimensione dello spazio comune.

Opzione	Descrizione
	Questo approccio prevede una rotazione ortogonale della configurazione comune, seguita da una proiezione su un sottospazio. In questo caso, la rotazione e la proiezione formano i cosiddetti pesi spaziali.

Condizionalità / Trasformazione

Questo riquadro controlla la trasformazione delle prossimità prima di inserire le distanze nella configurazione.

- Con l'opzione **Incondizionato**, è possibile applicare una singola funzione di trasformazione a tutte le fonti. Questa opzione è appropriata quando si presume che le fonti riflettano una struttura comune.

  La trasformazione predefinita è visualizzata in **Trasformazione** (ordinale con legami che possono essere sciolti).

  Fare clic su **Definisci trasformazione** per aprire una finestra di dialogo in cui è possibile scegliere tra tutti i tipi di trasformazione (lineare, potenza, monotonica, spline, ordinale), la gestione dei legami e altri parametri.
- Con l'opzione **Matrice-Condizionale** si possono scegliere diverse funzioni di trasformazione per diverse sorgenti o gruppi di sorgenti. Quando viene selezionata, ogni sorgente viene elencata con la funzione di trasformazione predefinita.

  Fare clic su **Definisci trasformazione** per scegliere un'altra funzione da applicare alle sorgenti selezionate. Questa opzione è abilitata solo quando è definita più di una sorgente.

In **Definisci trasformazione** è possibile selezionare il modo in cui le prossimità vengono trasformate in modo ottimale prima di essere abbinate alle distanze nella configurazione spaziale. Sono disponibili le seguenti opzioni:

Opzione	Descrizione
Nessuna	Non viene applicata alcuna trasformazione. Le prossimità sono utilizzate così come sono.
Lineare	Applica una trasformazione rettilinea alle vicinanze. **Intercetta** (casella di controllo): aggiunge una costante alla trasformazione, consentendo la modellazione su scala intervallare. Se non è selezionata, la trasformazione viene scalata per rapporto (non viene aggiunta alcuna costante).
Potenza	Applica una trasformazione di potenza. L' **intercetta** (casella di controllo) serve ad aggiungere uno spostamento costante alla funzione di potenza, se abilitata.
Spline	Adatta una funzione polinomiale piecewise (spline) liscia alle vicinanze. L' **intercetta** include un termine costante nella base della spline. Il **grado** imposta la complessità dei segmenti polinomiali (ad esempio, 1 = lineare, 2 = quadratico).

Opzione	Descrizione
	Nodi interni controlla il numero di punti interni utilizzati per definire la forma della spline.
Sequenza di nodi	Intervalli uguali: I nodi sono uniformemente distanziati lungo l'intervallo di prossimità. Percentili: I nodi sono posizionati in corrispondenza dei quantili empirici della distribuzione di prossimità.
Ordinale	Applica una trasformazione di rango (monotona). **Consentire lo scioglimento dei legami** : Le prossimità legate possono essere trasformate in valori diversi. **Mantenere i legami legati** : Le prossimità legate rimangono legate dopo la trasformazione.

Mappatura di prossimità: Restrizioni

Con la scheda **Restrizioni** è possibile limitare lo spazio delle soluzioni specificando le variabili degli attributi. Queste variabili guidano attivamente la configurazione, limitando lo spazio comune degli oggetti a una combinazione lineare degli attributi. Ciò consente una mappatura supervisionata, che migliora l'interpretabilità ancorando le dimensioni a variabili esterne significative.

Selezionare una delle seguenti opzioni.

Nessuno
Non vengono imposte restrizioni. La configurazione è determinata esclusivamente dalle prossimità (derivate).

Attributi
Abilita la mappatura supervisionata utilizzando una o più variabili esterne come vincoli di attributo. Selezionando questa opzione, è possibile definire quali variabili fungono da attributi e specificare come ciascuna di esse viene trasformata.

Variabili

L'elenco delle **variabili** a sinistra visualizza tutte le variabili disponibili nel set di dati. Selezionare una o più variabili e fare clic sul pulsante freccia per spostarle in **Attributi**.

Attributi

Nel riquadro di destra vengono visualizzate tutte le variabili di attributo selezionate e la funzione di trasformazione predefinita. Queste trasformazioni controllano il modo in cui gli attributi vengono scalati o trasformati per essere utilizzati nella configurazione ristretta.

Fare clic su **Definisci trasformazione** per assegnare una funzione di trasformazione all'attributo selezionato. Sono disponibili le seguenti opzioni:

- **Lineare** mantiene l'attributo sulla sua scala standardizzata.
- La **spline** applica una trasformazione liscia, monotonica o non monotonica, utilizzando una funzione polinomiale a passo variabile.
- L **'ordinale** conserva solo l'ordine di grado.
- **Nominal** tratta le categorie come non ordinate e le adatta in modo ottimale. Fornisce una singola trasformazione (a differenza della trasformazione nominale per le proprietà).

È possibile selezionare più righe nel campo **Attributi** per definire la stessa trasformazione per più attributi.

Lo spazio comune viene stimato in base al vincolo che le coordinate dell'oggetto si trovino all'interno del sottospazio coperto dagli attributi trasformati. Il campo **Numero attuale di dimensioni** visualizza la dimensionalità della soluzione, impostata nella scheda **Criteri**. Selezionare sempre un numero di attributi superiore al numero di dimensioni.

Mappatura di prossimità: Supplemento

Con la scheda **Supplementare** è possibile definire variabili di proprietà che aiutano l'interpretazione della configurazione spaziale ma non ne influenzano la stima (mappatura non supervisionata).

Variabili

L'elenco **Variabili** a sinistra visualizza tutte le variabili disponibili nel dataset attivo. Selezionare una o più variabili da trattare come proprietà e spostarle nell'elenco **Proprietà**. È possibile selezionare un numero qualsiasi di proprietà.

Proprietà

Il riquadro **Proprietà** elenca tutte le variabili designate come variabili di proprietà, insieme alla trasformazione assegnata a ciascuna di esse.

Fare clic su **Definisci trasformazione** per assegnare una funzione di trasformazione a una o più variabili di proprietà selezionate. Vedere le seguenti opzioni disponibili:

- **Linear** mantiene le proprietà sulla sua scala standardizzata.
- La **spline** applica una trasformazione liscia, monotonica o non monotonica, utilizzando una funzione polinomiale a tratti.
- L **'ordinale** conserva solo l'ordine di grado.
- Le categorie di fit **nominali** sono i centroidi dei punti oggetto associati nella configurazione.

Mappatura di prossimità: Criteri

La scheda **Criteri** specifica le impostazioni algoritmiche che controllano la dimensionalità della soluzione, il comportamento di convergenza del processo di ottimizzazione e la standardizzazione e l'orientamento post-soluzione della configurazione risultante.

Dimensionalità

Imposta il numero di dimensioni dello spazio comune. Il valore deve essere uguale o superiore a 2 e inferiore al valorenumero di oggetti - 1. L'aumento del numero di dimensioni può migliorare l'adattamento, ma può ridurre l'interpretabilità.

Convergenza dell'algoritmo

Queste impostazioni determinano il momento in cui l'ottimizzazione iterativa si interrompe.

Campo	Descrizione	Valore predefinito
Minimo della funzione	Imposta la soglia del valore della funzione di perdita al di sotto della quale la soluzione viene accettata. Se il valore della funzione di perdita scende al di sotto di questo minimo, l'iterazione si interrompe. Il valore deve essere superiore o uguale a zero.	$1.0 \times {}^{10\text{-}6}$

Campo	Descrizione	Valore predefinito
Differenza della funzione	Imposta il criterio di convergenza basato sulla variazione della funzione di perdita tra le iterazioni. Se la differenza assoluta tra i valori delle funzioni scende al di sotto di questo valore, si considera che l'algoritmo abbia raggiunto la convergenza. Il valore deve essere superiore o uguale a zero.	1.0×10^{-12}
Differenza della configurazione	Imposta la soglia di convergenza per le modifiche alla configurazione tra le iterazioni. Se il movimento totale delle posizioni degli oggetti scende al di sotto di questo valore, l'algoritmo si ferma. Il valore deve essere superiore o uguale a zero.	1.0×10^{-6}
Numero massimo di iterazioni	Specifica il numero massimo di iterazioni che l'algoritmo può eseguire. Se la convergenza non viene raggiunta entro questo numero, l'iterazione si interrompe comunque. Il valore deve essere superiore o uguale a zero.	1000

Standardizzazione

Specifica il modo in cui le variabili multivariate (se utilizzate per ricavare le prossimità) vengono standardizzate prima del calcolo della distanza. L'opzione si applica anche alle variabili di attributo e di proprietà.

- Per impostazione predefinita è selezionato **Somma dei quadrati = N**.
- Selezionare **Somma dei quadrati = N–1** se si desidera coerenza con altre procedure statistiche.

Dimensione

Specifica l'orientamento della configurazione dello spazio comune nello spazio delle coordinate.

Pulsante di scelta	Descrizione
Nessuno / Così com'è	Non vengono applicati vincoli. L'orientamento è quello degli assi principali, dove i segni delle dimensioni sono arbitrari.
Correlazione positiva tra dimensioni	Imposta la riflessione della configurazione per massimizzare la correlazione positiva tra le dimensioni e i dati.
Quadrante positivo per la massima coordinata assoluta	Riflette le dimensioni per garantire che l'oggetto con la coordinata più grande appaia nel quadrante positivo dello spazio.

Pulsante di scelta	Descrizione
Per dimensione	È possibile impostare le riflessioni individualmente per ogni dimensione. Selezionare la casella di controllo accanto a una quota per applicare la riflessione a quell'asse specifico.

Mappatura di prossimità: Uscita

Nella scheda **Output** è possibile richiedere tabelle e grafici che riassumono i dati e i risultati dell'analisi PROXMAP. Contiene sottotabelle per ciascuna delle principali schede di analisi (**Dati**, **Iniziale**, **Modello**, **Restrizioni**, **Supplementare** e **Criteri**)

È possibile richiedere un output relativo a qualsiasi aspetto dell'analisi. Ciascuna sottotabella contiene caselle di controllo per richiedere tabelle o grafici specifici.

Dati

Nella scheda **Dati** è possibile richiedere l'output relativo ai dati: dati multivariati, prossimità e pesi.

Tabelle (dati)

Opzione	Descrizione
Dati multivariati originali	Visualizza i valori grezzi delle variabili utilizzate per ricavare le prossimità.
Prossimità originali	Visualizza le prossimità fornite direttamente dall'utente (se applicabile).
Prossimità derivate	Mostra le prossimità calcolate dai dati multivariati per tutte le fonti.
Prossimità combinate	Visualizza le prossimità aggregate se si combinano più fonti.

Tabelle (pesi)

Opzione	Descrizione
Pesi originali	Elenca i pesi di prossimità grezzi (forniti dall'utente) utilizzati per ciascuna sorgente.
Pesi derivati	Fornisce elenchi di pesi derivati dai pesi grezzi, dai pesi dei casi oggetto, dai dati mancanti e dalla funzione di peso applicata alle prossimità.
Pesi degli oggetti (casi)	Visualizza i pesi applicati per oggetto (se specificati).

Grafici (dati)

L' **istogramma delle prossimità** traccia la distribuzione dei valori di prossimità tra tutte le fonti.

Grafici (pesi)

L' **istogramma dei pesi** traccia la distribuzione dei pesi di prossimità (se i pesi sono forniti o derivati).

Iniziale

Con la sottotabella **Iniziale**, è possibile richiedere l'output relativo alla configurazione iniziale, alla scomposizione dell'adattamento e alle misure diagnostiche dopo la stima iniziale.

Tabelle

Opzione	Descrizione
Coordinate	Visualizza le coordinate degli oggetti nella configurazione iniziale.
Direzioni variabili	Mostra le direzioni dei vettori delle variabili originali nello spazio di configurazione (per dati multivariati). Fornisce coordinate multiple per le variabili nominali (se presenti nei dati multivariati).
Direzioni delle proprietà	Visualizza le direzioni delle variabili di proprietà (abilitato solo quando le proprietà sono definite). Fornisce coordinate multiple per le variabili nominali (se presenti nei dati multivariati).
Decomposizione di stress	Riporta il contributo di ciascun oggetto al valore complessivo della perdita (sollecitazione).
diagnostica	Include statistiche di adattamento e diagnostica per la valutazione della soluzione iniziale. Questi includono autovalori e autovettori per il classico inizio di scala.

Grafici

Opzione	Descrizione
Configurazione	Visualizza una rappresentazione grafica dello spazio degli oggetti.
Biplot multivariato	Combina i punti dell'oggetto e le direzioni dei vettori variabili in un grafico congiunto, disponibile per i dati multivariati.
Biplot delle proprietà	Visualizza la relazione tra gli oggetti e le variabili delle proprietà (disponibile quando le proprietà sono definite).
Grafico scree	Grafica gli autovalori della soluzione scalare classica e le dimensioni rimanenti.
Grafico di stress	Un grafico a bolle sovrapposto allo spazio di configurazione dell'oggetto. Ogni oggetto è rappresentato come un punto circondato da un cerchio. La dimensione di ciascun cerchio riflette il contributo relativo di quell'oggetto alla sollecitazione normalizzata complessiva della configurazione.

Modello

Con la sottotabella **Modello** è possibile richiedere un output grafico e tabellare relativo agli spazi comuni e individuali, alle trasformazioni specifiche della fonte e alle misure di adattamento del modello.

Tabelle

Opzione	Descrizione
Coordinate	Elenca le coordinate degli oggetti nello spazio comune finale.
Direzioni variabili	Elenca le direzioni delle variabili originali (per i dati multivariati). Fornisce coordinate multiple per le variabili nominali (se presenti nei dati multivariati).
Pesi dello spazio	Visualizza i pesi spaziali stimati per ciascuna sorgente in base a modelli che ammettono differenze individuali.

Grafici (Spazio comune)

Opzione	Descrizione
Configurazione (selezionata per impostazione predefinita)	Visualizza la configurazione spaziale comune degli oggetti. **La configurazione** è un'impostazione predefinita inclusa nell'output, a meno che non si selezionino altre opzioni.
Biplot multivariato	Mostra che i dati di configurazione e di direzione sono multivariati.
Biplot del rango ridotto	Visualizza le proiezioni specifiche delle sorgenti nello spazio comune bidimensionale se il modello include spazi individuali di rango ridotto con rango 1. Questo grafico combina lo spazio comune con gli spazi individuali in un unico grafico.
Struttura al albero ricoprente il minimo	Collega tutti gli oggetti della configurazione con la lunghezza totale dei bordi più breve possibile. L'MST aiuta a visualizzare la struttura sottostante della configurazione, evidenziando i raggruppamenti e le transizioni basate sulla prossimità.
Vicini	Un grafo dei vicini collega ogni oggetto della configurazione ai suoi k vicini più prossimi, in base alle distanze nello spazio di configurazione. Questo grafico enfatizza la struttura locale e aiuta a rilevare cluster o outlier nella configurazione. È possibile scegliere il valore del numero di vicini più prossimi.
Soglia	Collega coppie di oggetti nello **spazio di configurazione** le cui vicinanze originali o trasformate sono inferiori o uguali a una soglia specificata. È possibile scegliere il valore della soglia.

Grafici (Spazio individuale)

Opzione	Descrizione
Configurazione	Traccia le configurazioni spaziali individuali per le sorgenti selezionate (usato con i modelli di differenze individuali).
Pesi dello spazio	Visualizza i pesi delle dimensioni per il modello di dilatazione e il modello euclideo ponderato.

Grafici (misure di trasformazione e adattamento)

Opzione	Descrizione
Trasformazione	Mostra le prossimità trasformate rispetto alle prossimità originali per le fonti selezionate.
Residui	Traccia le differenze tra le prossimità trasformate e le distanze adattate rispetto alle prossimità trasformate.
Adatta	Visualizza le distanze adattate rispetto alle distanze trasformate per sorgente.
Shepard	Crea un diagramma di Shepard che confronta le vicinanze e le distanze adattate rispetto alle vicinanze originali per le sorgenti selezionate.

Limitazioni

La sottotabella **Restrizioni** fornisce l'output relativo all'uso delle variabili di attributo come vincoli di configurazione (mappatura supervisionata). È possibile richiedere tabelle e visualizzazioni per le versioni originali, trasformate e adattate di questi attributi.

Attributi della tabella

Opzione	Descrizione
Attributi originali	Visualizza i valori grezzi delle variabili di attributo selezionate.
Attributi trasformati	Elenca i valori degli attributi dopo l'applicazione delle trasformazioni ottimali.
Pesi degli attributi	Visualizza i pesi calcolati per gli attributi per la restrizione dello spazio comune.
Direzioni degli attributi	Elenca le coordinate vettoriali associate a ciascun attributo nello spazio dei giunti.

Grafici

Selezionate uno o più attributi dalla seguente tabella per generare i seguenti grafici:

Opzione	Descrizione
Biplot	Visualizza la configurazione dell'oggetto con i vettori per le variabili di attributo selezionate.
Trasformazione	Visualizza la funzione di trasformazione applicata a ciascun attributo selezionato.

Opzione	Descrizione
Residui	Visualizza la differenza tra i valori trasformati adattati e i valori trasformati per gli attributi selezionati.

Il **grafico delle proiezioni** visualizza una proiezione unidimensionale degli oggetti nello spazio comune su ciascuno degli attributi separatamente.

Supplementare

La sottotabella **Supplementare** fornisce l'output relativo alle variabili di proprietà, che vengono utilizzate per l'interpretazione della configurazione ma non influiscono sull'ottimizzazione (mappatura non supervisionata).

Proprietà delle tabelle

Opzione	Descrizione
Proprietà originali	Visualizza i valori grezzi delle variabili di proprietà.
Proprietà trasformate	Mostra i valori dopo l'applicazione delle trasformazioni ottimali.
Direzioni delle proprietà	Elenca i vettori direzionali delle proprietà nello spazio di configurazione (se applicabile).

Proprietà dei grafici

Selezionare una o più proprietà dall'elenco per generare i seguenti grafici:

Opzione	Descrizione
Biplot	Visualizza la configurazione dell'oggetto con i vettori per le variabili di proprietà selezionate.
Trasformazione	Mostra la trasformazione di ciascuna proprietà selezionata.
Residui	Traccia la differenza tra i valori trasformati e adattati e i valori trasformati per le proprietà selezionate.

Il **grafico delle proiezioni** visualizza una proiezione unidimensionale degli oggetti nello spazio comune su ciascuna proprietà separatamente.

Criteri

La scheda **Criteri** fornisce informazioni diagnostiche e visualizzazioni relative al processo di stima iterativa e alla convergenza del modello.

Tabelle

Opzione	Descrizione
Riepilogo (selezionato per impostazione predefinita)	Riporta i valori finali di sollecitazione, il numero di iterazioni, il tempo trascorso ed eventuali messaggi di errore. **Il riepilogo** è un'impostazione predefinita inclusa nell'output, a meno che non si selezionino altre opzioni.

Opzione	Descrizione
diagnostica	Elenca ulteriori statistiche di adattamento, tra cui la formula di Kruskal per lo stress 1, il coefficiente di alienazione, la dispersione contabilizzata, il coefficiente di congruenza di Tucker, le correlazioni di Pearson e Spearman.
Cronologia delle iterazioni	Visualizza la sollecitazione normalizzata per ogni iterazione, la differenza di sollecitazione tra le iterazioni e la variazione di configurazione (differenza di coordinate) tra le iterazioni. Per limitare la dimensione della tabella, la tabella della cronologia delle iterazioni riporta solo un sottoinsieme di iterazioni, seguendo le potenze di 2. Questo cattura il modello di convergenza generale senza elencare tutte le iterazioni
Decomposizione di stress	Scompone il valore complessivo della sollecitazione nei contributi dei singoli oggetti.

Grafici

Opzione	Descrizione
Cronologia delle iterazioni	Visualizza il grafico di un sottoinsieme dei valori per le iterazioni, seguendo potenze di 2. Valori delle sollecitazioni normalizzate in funzione del numero di iterazioni.
Grafico di stress	Un grafico a bolle che mostra i contributi degli oggetti nello spazio comune visualizzando cerchi intorno ai punti degli oggetti che sono proporzionali al loro contributo alla sollecitazione normalizzata.

Mappatura di prossimità: Salva

Con la scheda **Salva** è possibile esportare i risultati selezionati in un nuovo set di dati.

È possibile scegliere di salvare i seguenti elementi:

Coordinate iniziali
Le coordinate della configurazione iniziale (spazio iniziale).

Coordinate finali
Le coordinate della configurazione finale (spazio comune).

Coordinate del modello
Le coordinate delle singole configurazioni (spazi individuali). L'opzione è attivata solo se il numero di sorgenti è superiore a uno.

Prossimità derivate
Le prossimità derivate dai dati multivariati o le prossimità originali se fornite direttamente.

Prossimità combinate
Le prossimità che sono aggregate su tutte le fonti (se sono definite più fonti).

Prossimità trasformate
Le prossimità dopo aver applicato le funzioni di trasformazione selezionate.

Attributi trasformati
Gli attributi trasformati. L'opzione è abilitata solo se gli attributi sono specificati nella scheda **Restrizioni**.

Attributi adattati
I valori degli attributi adattati alla configurazione. L'opzione è abilitata solo se sono selezionati gli attributi.

Proprietà trasformate
I valori trasformati delle proprietà supplementari. L'opzione è abilitata solo se le proprietà sono definite nella scheda **Supplementare**.

Proprietà adattate
I valori delle proprietà proiettati sulla configurazione. L'opzione è abilitata solo se sono selezionate le **Proprietà**.

Scegli dataset

Gli elementi selezionati vengono scritti in un nuovo set di dati. Fare clic sul pulsante ellissi (...) per aprire la finestra di dialogo **Salva dati con nome**, dove è possibile specificare il nome del file e la cartella di destinazione del nuovo set di dati.

Correlazioni parziali

La procedura Correlazioni parziali consente di calcolare i coefficienti di correlazione parziale che descrivono la relazione lineare tra due variabili controllando gli effetti di una o più variabili aggiuntive. Le correlazioni sono misure di associazione lineare. Due variabili possono essere perfettamente correlate, ma se la relazione non è lineare, un coefficiente di correlazione non rappresenta la statistica più adatta a misurarne l'associazione.

Esempio
Esiste una correlazione tra i fondi stanziati per il sistema sanitario e il tasso di malattie? Sebbene ci si potrebbe aspettare che una tale relazione sia di tipo negativo, uno studio riporta una correlazione *positiva* significativa: mentre aumenta il finanziamento per l'assistenza sanitaria, sembrano aumentare i tassi di malattia. Il controllo della frequenza delle visite ai medici elimina virtualmente la correlazione positiva osservata. I fondi stanziati per il sistema sanitario e il tasso di malattie sembrano avere una correlazione positiva solo perché più pazienti hanno accesso al sistema sanitario quando vengono aumentati i fondi, con un conseguente aumento delle malattie segnalate da medici e ospedali.

Statistiche
Per ogni variabile: numero di casi con valori non mancanti, media e deviazione standard. Matrici di correlazione parziale e di ordine zero, con gradi di libertà e livelli di significatività.

Considerazioni sui dati

Dati
Utilizzare variabili simmetriche, quantitative.

Ipotesi
La procedura Correlazioni parziali si fonda sull'ipotesi che ciascuna coppia di variabili sia bivariata normale.

Ottenere Correlazioni Parziali

1. Dai menu, scegliere:

   Analizza > **Correlazione** > **Parziale...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare due o più variabili numeriche per cui è necessario calcolare le correlazioni parziali.
3. Selezionare una o più variabili numeriche di controllo.

Sono inoltre disponibili le seguenti opzioni:

Test di significatività
È possibile selezionare le probabilità a due code o a una tazzina. Se si conosce in anticipo la direzione dell'associazione, selezionare **A una coda.**. In alternativa, selezionare **A due code**.

Visualizza il livello di significatività effettivo
Per impostazione predefinita, vengono indicati la probabilità e i gradi di libertà di ciascun coefficiente di correlazione. Se questa opzione viene deselezionata, i coefficienti significativi al livello 0,05 vengono identificati con un solo asterisco, i coefficienti significativi al livello 0,01 con un doppio asterisco e i gradi di libertà vengono eliminati. Questa impostazione viene applicata alle matrici di correlazione parziale e di ordine zero.

Correlazioni parziali: Opzioni

Statistiche. È possibile scegliere una delle seguenti opzioni o entrambe:

- **Medie e deviazioni standard.** Visualizzate per ciascuna variabile. Viene indicato anche il numero di casi con valori non mancanti.
- **Correlazioni di ordine zero.** Viene visualizzata una matrice di correlazioni semplici tra tutte le variabili, incluse le variabili di controllo.

Valori mancanti. È possibile scegliere una delle seguenti opzioni:

- **Escludi casi listwise.** I casi con valori mancanti per qualsiasi variabile, incluse le variabili di controllo, vengono esclusi da tutti i calcoli.
- **Escludi casi pairwise.** Per il calcolo delle correlazioni di ordine zero su cui si basano le correlazioni parziali, i casi con valori mancanti per una o entrambe le variabili di una coppia non verranno utilizzati. L'eliminazione pairwise consente di utilizzare il massimo numero di dati possibile. Il numero di casi, tuttavia, può differire a seconda del coefficiente. Quando è attiva l'eliminazione pairwise, i gradi di libertà per un particolare coefficiente parziale si basano sul numero minimo di casi utilizzati per il calcolo di una delle correlazioni di ordine zero.

Funzioni aggiuntive del comando PARTIAL CORR

Il linguaggio della sintassi dei comandi consente inoltre di:

- Leggere una matrice di correlazione di ordine zero o scrivere una matrice di correlazione parziale (con il sottocomando `MATRIX`).
- Ottenere correlazioni parziali tra due elenchi di variabili (con la parola chiave `WITH` nel sottocomando `VARIABLES`).
- Ottenere più analisi (con più sottocomandi `VARIABLES`).
- Specificare i valori degli ordini da richiedere (ad esempio sia le correlazioni parziali di primo e secondo ordine) quando sono disponibili due variabili di controllo (con il sottocomando `VARIABLES`).
- Eliminare i coefficienti ridondanti (con il sottocomando `FORMAT`).
- Visualizzare una matrice di correlazioni semplici quando non è possibile calcolare alcuni coefficienti (con il sottocomando `STATISTICS`).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Distanze

Questa procedura consente di calcolare una grande varietà di statistiche in base alle similarità o alle dissimilarità (distanze), sia considerando coppie di variabili, sia considerando coppie di casi. Tali misure della distanza o similarità possono poi essere utilizzate insieme ad altre procedure quali analisi fattoriale, analisi cluster o scaling multidimensionale per l'analisi di dataset complessi.

Esempio. È possibile misurare le similarità tra coppie di automobili in base a caratteristiche specifiche, quali dimensione del motore, MPG e potenza. Grazie al calcolo delle similarità, è possibile stabilire quali automobili sono simili e quali sono differenti tra loro. Per un'analisi più formale, si può scegliere

di applicare alle similarità l'analisi cluster gerarchica o lo scaling multidimensionale che consentono di esplorare la struttura sottostante.

Statistiche. Le misure di dissimilarità (distanza) disponibili per i dati di intervallo sono: distanza euclidea, distanza euclidea al quadrato, Chebychev, City-block, Minkowski, personalizzata. Per i dati di conteggio, le misure disponibili sono chi-quadrato o phi-quadrato. Per i dati binari, infine, le misure disponibili sono: distanza euclidea, distanza euclidea al quadrato, differenza di dimensione, differenza di modello, varianza, forma, Lance e Williams. Le misure di similarità disponibili per i dati di intervallo sono la correlazione di Pearson o il coseno. Per i dati binari sono invece disponibili le seguenti misure: Russel e Rao, corrispondenza semplice, Jaccard, Dice, Rogers e Tanimoto, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Kulczynski 1, Kulczynski 2, Sokal e Sneath 4, Hamann, Lambda, *D* di Anderberg, *Y* di Yule, *Q* di Yule, Ochiai, Sokal e Sneath 5, correlazione phi a 4 punti o dispersione.

Per ottenere matrici delle distanze

1. Dai menu, scegliere:

   Analizza > **Correlazione** > **Distanze...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una variabile numerica per calcolare la distanza tra casi e almeno due variabili numeriche per calcolare la distanza tra variabili.
3. Scegliere un'alternativa nel gruppo Calcola distanze per calcolare le prossimità tra casi o tra variabili.

Distanze: Misure di dissimilarità

Dal gruppo Misura selezionare l'alternativa corrispondente al tipo di dati desiderato (intervallo, conteggio o binari), quindi scegliere dall'elenco a discesa una delle misure corrispondenti a quel tipo di dati. Le misure disponibili per tipo di dati sono le seguenti:

- **Dati di intervallo.** Distanza euclidea, distanza euclidea al quadrato, Chebychev, City-block, Minkowski o personalizzata.
- **Dati di conteggio.** Misura chi-quadrato e misura phi-quadrato.
- **Dati binari.** Distanza euclidea, distanza euclidea al quadrato, differenza di dimensione, differenza di modello, varianza, forma o di Lance e Williams. (Inserire i valori Presente e Assente per specificare i due valori significativi, tutti gli altri valori verranno ignorati dalle distanze).

Il gruppo Trasforma valori consente di standardizzare i valori dei dati per casi o valori *prima* di calcolare le prossimità. Tali trasformazioni non sono applicabili ai dati binari. I metodi di standardizzazione disponibili sono punteggi *z*, intervallo da –1 a 1, intervallo da 0 a 1, grandezza massima di 1, media di 1 o deviazione standard di 1.

Il gruppo Trasforma misure consente di trasformare i valori generati dalla misura della distanza. Questi verranno applicati dopo il calcolo della misura della distanza. Le opzioni disponibili sono Valori assoluti, Cambia segno e Modifica scala su intervallo 0-1.

Distanze: Misure di similarità

Selezionare l'alternativa corrispondente al tipo di dati desiderato (intervallo o binari) dal gruppo Misura , quindi scegliere una delle misure corrispondenti a quel tipo di dati dall'elenco a discesa. Le misure disponibili per tipo di dati sono le seguenti:

- **Dati di intervallo.** Correlazione Pearson o coseno.
- **Dati binari.** Russel e Rao, corrispondenza semplice, Jaccard, Dice, Rogers e Tanimoto, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Kulczynski 1, Kulczynski 2, Sokal e Sneath 4, Hamann, Lambda, *D* di Anderberg, *Y* di Yule, *Q* di Yule, Ochiai, Sokal e Sneath 5, correlazione phi a 4 punti o dispersione. (Inserire i valori Presente e Assente per specificare i due valori significativi, tutti gli altri valori verranno ignorati dalle distanze).

Il gruppo Trasforma valori consente di standardizzare i valori dei dati per casi o valori prima di calcolare le prossimità. Tali trasformazioni non sono applicabili ai dati binari. I metodi di standardizzazione disponibili sono punteggi *z*, intervallo da –1 a 1, intervallo da 0 a 1, grandezza massima di 1, media di 1 e deviazione standard di 1.

Il gruppo Trasforma misure consente di trasformare i valori generati dalla misura della distanza. Questi verranno applicati dopo il calcolo della misura della distanza. Le opzioni disponibili sono Valori assoluti, Cambia segno e Modifica scala su intervallo 0-1.

Funzioni aggiuntive del comando PROXIMITIES

La procedura Distanze usa la sintassi del comando PROXIMITIES. Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare un intero come potenza per la misura della distanza di Minkowski.
- Specificare qualsiasi intero come potenza e radice per la misura della distanza personalizzata.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Correlazione della distanza

Questa funzione richiede il modulo Statistics Base.

La procedura delle correlazioni a distanza calcola la dipendenza tra due variabili casuali o dati multivariati. Valuta la relazione non lineare delle variabili, a differenza dei metodi di correlazione tradizionali.

Esempio

Un'azienda di vendita al dettaglio vuole capire se i dati demografici dei clienti (età, reddito, istruzione, stato civile) influenzano il loro comportamento d'acquisto. I metodi di correlazione tradizionali possono fallire se le relazioni sono non lineari o comportano interazioni tra più fattori.

La procedura di correlazione a distanza può essere applicata nello scenario con i seguenti vantaggi:

- Cattura le dipendenze non lineari. A differenza della correlazione di Pearson, la correlazione a distanza rileva le relazioni anche se non sono lineari.
- Gestisce relazioni multidimensionali. Può valutare come una combinazione di fattori demografici sia correlata ai modelli di acquisto.
- Fornisce una misura completa. Una correlazione di distanza significativa suggerisce che i dati demografici dei clienti influenzano fortemente il comportamento d'acquisto, guidando strategie di marketing mirate.

Statistiche associate

- Correlazione a distanza (*dCor(X,Y)): Misura la dipendenza tra i dati demografici del cliente (X*) e il comportamento d'acquisto (*Y*).
- Valore p basato sulle permutazioni: Valuta la significatività statistica della dipendenza. Un piccolo valore p (ad esempio, p<0.05) indica una forte relazione.
- Covarianza a distanza (*dCov(X,Y)): Quantifica la forza della dipendenza tra le variabili.*
- Distanza Varianza (*dVar(X), dVar(Y)*): misura la variabilità all'interno delle singole variabili per normalizzare la correlazione.

Considerazioni sui dati

Solo le variabili continue possono essere utilizzate per l'analisi. Per calcolare la correlazione di distanza è necessario un minimo di 2 variabili continue. È consentito un massimo di 30 variabili.

Ottenere la correlazione di distanza

1. Dal menu, fare clic su: **Analizza** > **Correlazione Distanza** > **Correlazione**.

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare due o più variabili numeriche per le quali calcolare la correlazione di distanza. È possibile selezionare fino a 30 variabili continue dal set di dati di origine per l'analisi.
3. Selezionare una variabile identificativa. Se si opta per una **variabile ID**, i risultati visualizzano una tabella a matrice di distanza con gli ID.
4. Fare clic su **OK** per eseguire la procedura con le impostazioni specificate e generare l'output.

Correlazione della distanza: Criteri

La sottodialogo Criteri fornisce i parametri di controllo per la normalizzazione e altre opzioni chiave per personalizzare la procedura di correlazione delle distanze.

Normalizzazione

Selezionare l'opzione per assegnare un tipo di normalizzazione al set di dati. La normalizzazione standardizza le scale delle variabili, ma potrebbe inavvertitamente distorcere le relazioni altamente non lineari nei dati, che sono cruciali per valutare accuratamente la correlazione delle distanze. Quindi, è stato incluso per facilitare le relazioni lineari.

- La normalizzazione **Min-Max** scala ogni caratteristica a un intervallo fisso, di solito [0, 1]. Questa trasformazione è utile quando il set di dati contiene variabili con unità o intervalli diversi. Questa normalizzazione consente un confronto equo tra le variabili, assicurando che nessuna singola variabile domini i calcoli della distanza a causa della sua scala. Aiuta a mitigare gli effetti di grandi intervalli numerici.
- La normalizzazione dello **Z-Score** centra i dati intorno alla media con una deviazione standard di 1. Questo approccio è utile per gli insiemi di dati che sono normalmente distribuiti. Questa normalizzazione riduce l'influenza degli outlier, in quanto conferisce a ciascuna variabile una scala standardizzata. Consente una più facile interpretazione dei risultati, soprattutto quando si confrontano variabili diverse, poiché tutte le variabili sono trattate allo stesso modo rispetto alla loro distribuzione.
- **Robust-Scaling** utilizza la mediana e l'intervallo interquartile (IQR) per scalare i dati, rendendoli resistenti agli outlier. L'influenza degli outlier è ridotta al minimo, consentendo di calcolare in modo più affidabile la correlazione delle distanze in presenza di valori estremi. Questa tecnica è utile quando il set di dati contiene anomalie che potrebbero falsare i risultati.
- La **Trasformazione Log** è progettata per ridurre la skewness dei dati, in particolare per le distribuzioni con skewed positivo. Questo può aiutare a stabilizzare la varianza nel set di dati. La trasformazione logica comprime i valori grandi ed espande quelli piccoli, contribuendo a normalizzare la distribuzione. Ciò può migliorare l'accuratezza dell'analisi della correlazione delle distanze, soprattutto quando si ha a che fare con insiemi di dati altamente distorti.

Intervallo di confidenza Percentuale

Specificare la percentuale dell'intervallo di confidenza da utilizzare nell'analisi statistica.

La percentuale dell'intervallo di confidenza indica il livello di certezza associato alle stime statistiche. Una percentuale più alta riflette una maggiore fiducia nell'accuratezza dell'intervallo, ma si traduce in un intervallo più ampio, mentre una percentuale più bassa produce un intervallo più ristretto con una minore certezza

Permutazioni di correlazione della distanza

Nella correlazione a distanza, il test di permutazione viene utilizzato per valutare la significatività statistica della correlazione a distanza osservata tra due variabili. L'obiettivo è valutare se la correlazione osservata è significativamente diversa da quella che potrebbe verificarsi per caso. Una delle serie di dati (o entrambe) viene permutata in modo casuale, interrompendo qualsiasi associazione esistente tra le variabili. La correlazione di distanza viene quindi ricalcolata per ogni permutazione, generando una distribuzione dei valori di correlazione di distanza sotto l'ipotesi nulla di assenza di associazione.

Selezionare **Imposta seme personalizzato** per abilitare il test di permutazione e specificare il numero di iterazioni e il livello di significatività. Impostare il valore in **Semi**. È possibile impostare

facoltativamente un seme per il generatore di numeri casuali utilizzato nei test di permutazione. Il seme deve essere un numero intero positivo o zero. Il valore predefinito è 2000000.

Impostare le **Iterazioni massime**. Il valore deve essere un intero positivo. Il valore predefinito è impostato su 100 e non sono ammessi valori negativi.

Confrontando la correlazione di distanza osservata con la distribuzione delle correlazioni di distanza da insiemi di dati permutati, si calcola un valore p. Il valore p rappresenta la percentuale di correlazioni di distanza permutate che sono maggiori o uguali alla correlazione di distanza osservata. Il livello di significatività viene utilizzato come soglia per determinare la significatività statistica. Se il valore p è inferiore al livello di significatività, la correlazione osservata è considerata statisticamente significativa. Impostare il valore del **Livello di significatività** tra 0 e 1. Il valore predefinito è 0.05.

Correlazione della distanza: Stampa

Stampa dettagli
Selezionare **Stampa dettagli** per visualizzare una tabella con le seguenti misure essenziali: numero di osservazioni (N), media, deviazione standard (Std. Deviazione), e i valori minimi e massimi per ogni variabile. Queste statistiche forniscono una panoramica completa del set di dati e consentono di comprendere meglio le distribuzioni e le caratteristiche delle variabili confrontate.

Coefficienti di correlazione della distanza
Selezionare **Coefficienti di correlazione a distanza** per generare una tabella che mostra i coefficienti di correlazione a distanza calcolati per l'analisi. `dCor Value` mostra il coefficiente di correlazione a distanza tra le variabili. Se si sceglie di **impostare un seme personalizzato** nella finestra di dialogo secondaria **Criteri** e si imposta un livello di significatività, è possibile visualizzare la colonna `Sig.` che indica il livello di significatività (valore p) associato al coefficiente di correlazione calcolato.

`Number of Replicates (R)` indica il numero di repliche (o permutazioni, se optato) utilizzate nell'analisi.

Stime di covarianza della distanza
Se si seleziona un livello di intervallo di confidenza (IC), la tabella include i limiti superiore e inferiore degli intervalli di confidenza, fornendo una visione più completa delle stime.

Metriche di distanza e stime di significatività
Le stime aiutano a valutare le relazioni e le dipendenze tra le variabili analizzate.

Matrice di distanza
Selezionare **Matrice di distanza** per generare una tabella che visualizzi la matrice di distanza per i dati monodimensionali (ad esempio. variabile "Età"). Vengono rappresentate le distanze tra ogni coppia di osservazioni. Gli elementi diagonali mostrano una distanza nulla (distanza da sé). Questa matrice fornisce una visione chiara delle distanze a coppie, in modo da poter effettuare un'analisi completa delle relazioni tra le variabili. Selezionare l'opzione per configurare una matrice di distanza. Nella finestra di dialogo **Configura matrice di distanza**, selezionare le variabili dall'elenco per configurare la matrice di distanza.

Salvare la matrice di distanza come file
Selezionare l'opzione per salvare le matrici di distanza in un file. Fornire il nome del file in **Outfile File Directory**. L'estensione del file supportata è `.sav`.

Correlazione della distanza: grafico

Selezionare **Plot** per generare i grafici per la Correlazione di distanza. Selezionare **Scatterplot** per generare un diagramma di dispersione che mostra le distanze o le correlazioni a coppie tra le variabili selezionate.

Selezionare la variabile dall'elenco per tracciarla sull'asse X. La variabile deve essere mutuamente esclusiva con le altre variabili scelte per lo scatterplot.

Selezionate un'altra variabile dall'elenco per tracciarla sull'asse Y. Questa variabile deve essere mutuamente esclusiva con le altre variabili scelte per lo scatterplot.

Modelli lineari

I modelli lineari eseguono la previsione per un obiettivo continuo in base alle relazioni lineari tra l'obiettivo e uno o più predittori.

I modelli lineari sono relativamente semplici e offrono una formula matematica di facile interpretazione per il calcolo del punteggio. Le proprietà di questi modelli sono ben note; in genere possono essere creati molto rapidamente rispetto agli altri tipi di modelli (ad esempio le reti neurali o gli alberi decisionali) nello stesso dataset.

Esempio. Una compagnia di assicurazioni con risorse limitate per indagare sulle richieste di indennizzo di proprietari di case desidera creare un modello per la stima dei costi delle richieste. Con il deployment di questo modello nei centri servizi, i rappresentanti possono inserire gli estremi delle richieste di risarcimento mentre sono al telefono con i clienti e ottenere immediatamente il costo "previsto" del risarcimento in base ai dati passati.

Requisiti del campo. Devono essere presenti un obiettivo ed almeno un input. Per impostazione predefinita, i campi con i ruoli predefiniti Entrambi e Nessuno non vengono utilizzati.

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

L'obiettivo deve essere continuo (scala). Non vi sono limitazioni del livello di misurazione sui predittori (input); i campi categoriali (nominale e ordinale) vengono utilizzati come fattori nel modello e i campi continui vengono utilizzati come covariate.

Nota: se un campo categoriale contiene più di 1000 categorie, la procedura non viene eseguita e non viene creato alcun modello.

Come ottenere un modello lineare

Questa funzione richiede il modulo Statistics Base.

Dai menu, scegliere:

Analizzare > **Regressione** > **Modelli lineari automatici ...**

1. Verificare che vi sia almeno un obiettivo e un input.
2. Fare clic su **Opzioni di creazione** per specificare le impostazioni facoltative del modello e di creazione.
3. Fare clic su **Opzioni modello** per salvare i punteggi nel dataset attivo ed esportare il modello in un file esterno.
4. Fare clic su **Esegui** per eseguire la procedura e creare gli oggetti Modello.

Obiettivi

Qual è l'obiettivo principale? Selezionare l'obiettivo appropriato.

- **Crea un modello standard.** Il metodo crea un singolo modello per prevedere l'obiettivo utilizzando i predittori. In genere i modelli standard sono più semplici da interpretare e il calcolo del punteggio può risultare più rapido rispetto agli insiemi di cui è stato eseguito il boosting, il bagging o agli insiemi di dataset di grandi dimensioni.
- **Migliora la precisione del modello (boosting).** Il metodo crea un modello dell'insieme tramite boosting, che genera una sequenza di modelli per ottenere previsioni più precise. La creazione e il calcolo del punteggio degli insiemi può richiedere più tempo rispetto ai modelli standard.

  Il boosting produce una successione di "modelli di componente", ognuno dei quali viene costruito sul dataset intero. Prima di costruire ogni modello di componente successivo, i record vengono pesati sulla base dei residui del modello di componente precedente. Ai casi con residui grandi vengono assegnati pesi dell'analisi relativamente più elevati, in modo che il modello del componente successivo si concentrerà sulla previsione corretta di tali record. Insieme questi modelli del componente formano un modello dell'insieme. Il modello dell'insieme calcola il punteggio di nuovi record utilizzando una regola di combinazione; le regole disponibili dipendono dal livello di misurazione dell'obiettivo.

- **Migliora la stabilità del modello (bagging).** Il metodo crea un modello dell'insieme tramite bagging (aggregazione bootstrap), che genera più modelli per ottenere previsioni più affidabili. La creazione e il calcolo del punteggio degli insiemi può richiedere più tempo rispetto ai modelli standard.

  L'aggregazione di bootstrap (bagging) genera delle repliche del dataset di addestramento, eseguendo il campionamento con sostituzione dal dataset originale. Ciò crea dei campioni di bootstrap di uguale dimensione nel dataset originale. Quindi, un "modello del componente" viene creato in ciascuna replica. Insieme questi modelli del componente formano un modello dell'insieme. Il modello dell'insieme calcola il punteggio di nuovi record utilizzando una regola di combinazione; le regole disponibili dipendono dal livello di misurazione dell'obiettivo.
- **Creare un modello per dataset di grandi dimensioni (richiede IBM SPSS Statistica Server).** Il metodo crea un modello dell'insieme suddividendo il dataset in blocchi di dati. Scegliere questa opzione se il dataset è troppo grande per creare uno dei modelli descritti oppure per la creazione di modelli incrementali. Questa opzione risulta più rapida per la creazione, ma può richiedere più tempo per il calcolo del punteggio rispetto a un modello standard. Questa opzione richiede la connettività IBM SPSS Statistica Server.

Consultare “Insiemi ” a pagina 164 per le impostazioni relative a boost, bagging e datasets molto grandiboost e bagging.

Di base

Prepara dati automaticamente. Questa opzione consente alla procedura di trasformare internamente l'obiettivo e i predittori per espandere al massimo la potenza predittiva del modello; le trasformazioni vengono salvate con il modello e applicate ai nuovi dati per il calcolo del punteggio. Le versioni originali dei campi trasformati vengono escluse dal modello. Per impostazione predefinita, viene eseguita le preparazioni automatica dei dati riportata di seguito.

- **Gestione di data e ora.** Ciascun predittore della data viene trasformato in un nuovo predittore continuo contenente il tempo trascorso da una data di riferimento (1970-01-01). Ogni predittore dell'ora viene trasformato in un nuovo predittore continuo contenente l'ora trascorsa da un'ora di riferimento (00:00:00).
- **Adatta livello di misurazione.** I predittori continui con meno di 5 valori distinti vengono riformulati come predittori ordinali. I predittori ordinali con più di 10 valori distinti vengono riformulati come predittori continui.
- **Gestione dei valori anomali.** I valori dei predittori continui che si trovano oltre un valore di interruzione (3 deviazioni standard dalla media) vengono impostati sul valore di interruzione.
- **Gestione valori mancanti.** I valori mancanti dei predittori nominali vengono sostituiti con la modalità della partizione di addestramento. I valori mancanti dei predittori ordinali vengono sostituiti dalla mediana della partizione di addestramento. I valori mancanti dei predittori continui vengono sostituiti dalla media della partizione di addestramento.
- **Unione supervisionata.** Ciò rende più parsimonioso un modello, riducendo il numero di campi da elaborare insieme all'obiettivo. Le categorie simili vengono identificate in base alla relazione tra input e obiettivo. Le categorie che non sono significativamente differenti (ovvero, che hanno un valore P maggiore di 0.1) vengono unite. Se tutte le categorie vengono unite in un'unica categoria, le versioni originali e derivate del campo vengono escluse dal modello perché non hanno alcun valore come predittore.

Livello di confidenza. Si tratta del livello di confidenza utilizzato per calcolare le stime dell'intervallo dei coefficienti del modello nella vista Coefficienti. Specificare un valore maggiore di 0 e minore di 100. Il valore predefinito è 95.

Selezione modello

Metodo di selezione del modello. Scegliere uno dei metodi di selezione del modello (dettagli di seguito) o **Includi tutti i predittori**, che semplicemente immette tutti i predittori disponibili come termini del modello di effetti principali. Per impostazione predefinita. viene utilizzata l'opzione **Stepwise in avanti**.

Selezione stepwise in avanti. All'inizio non vi sono effetti nel modello e questi vengono aggiunti e rimossi una fase per volta finché non possono più essere aggiunti o rimossi in base ai criteri stepwise.

- **Criteri per immissione/rimozione.** Si tratta della statistica utilizzata per determinare se un effetto deve essere aggiunto o rimosso nel modello. **Criterio di informazione (AICC)** si basa sulla verosimiglianza dell'insieme di addestramento, dato il modello, e viene adattato per penalizzare i modelli eccessivamente complessi. **Statistiche F** si basano su un test statistico del miglioramento nell'errore del modello. **R-quadrato adattato** si basa sull'adattamento dell'insieme di addestramento e viene adattato per penalizzare i modelli eccessivamente complessi. **Criterio di prevenzione del sovradattamento (ASE)** si basa sull'adattamento (ASE, average squared error) dell'insieme di prevenzione sovradattato. L'insieme di prevenzione sovradattato è un sottocampione casuale di circa il 30% del dataset originale che non è utilizzato per formare il modello.

  Se si sceglie un criterio diverso da **Statistiche F**, in ogni fase l'effetto corrispondente all'incremento positivo più grande nel criterio viene aggiunto al modello. Gli effetti nel modello che corrispondono a un decremento nel criterio vengono rimossi.

  Se si sceglie **Statistiche F** come criterio, in ogni fase l'effetto che ha il valore *p* più piccolo inferiore alla soglia specificata, **Includi effetti con valori p inferiori a**, viene aggiunto al modello. Il valore predefinito è 0.05. Gli effetti nel modello con un valore *p* superiore alla soglia specificata, **Rimuovi effetti con valori p maggiori di**, vengono rimossi. Il valore predefinito è 0.10.
- **Personalizza numero massimo di effetti nel modello finale.** Per impostazione predefinita, tutti gli effetti disponibili possono essere immessi nel modello. In alternativa, se l'algoritmo stepwise termina una fase con il numero massimo di effetti specificato, l'algoritmo termina con l'insieme corrente di effetti.
- **Personalizza numero massimo di fasi.** L'algoritmo stepwise termina dopo un determinato numero di fasi. Per impostazione predefinita, è tre volte il numero di effetti disponibili. In alternativa, specificare un numero intero positivo come numero massimo di fasi.

Selezione sottoinsiemi migliori. Controlla "tutti i possibili" modelli o almeno un sottoinsieme di modelli possibili più grande di quello stepwise in avanti per scegliere il migliore in base al criterio Sottoinsiemi migliori. **Criterio di informazione (AICC)** si basa sulla probabilità che il set di formazione abbia dato il modello, ed è regolato per penalizzare modelli troppo complessi. **R-quadrato adattato** si basa sull'adattamento dell'insieme di addestramento e viene adattato per penalizzare i modelli eccessivamente complessi. **Criterio di prevenzione del sovradattamento (ASE)** si basa sull'adattamento (ASE, average squared error) dell'insieme di prevenzione sovradattato. L'insieme di prevenzione sovradattato è un sottocampione casuale di circa il 30% del dataset originale che non è utilizzato per formare il modello.

Il modello con il valore più grande del criterio viene scelto come modello migliore.

Nota: La selezione dei sottoinsiemi migliori è maggiormente impegnativa a livello di calcolo rispetto alla selezione stepwise in avanti. Quando si utilizza la selezione Sottoinsiemi migliori in combinazione con attività di boosting, bagging o di dataset molto grandi, la creazione può richiedere molto più tempo di un modello standard creato utilizzando la selezione stepwise in avanti.

Insiemi

Queste impostazioni determinano il funzionamento della procedura di insieme che si verifica in caso di richiesta di boosting, bagging o dataset di grandi dimensioni negli obiettivi. Le opzioni che non si applicano all'obiettivo selezionato vengono ignorate.

Bagging e dataset di grandi dimensioni. Quando si calcola il punteggio di un insieme, questa è la regola utilizzata per combinare i valori previsti dai modelli di base per calcolare il valore del punteggio dell'insieme.

- **Regola di combinazione predefinita per gli obiettivi continui.** I valori previsti dell'insieme per gli obiettivi continui possono essere combinati utilizzando la media o la mediana dei valori previsti ricavati dai modelli di base.

Si osservi che quando l'obiettivo consiste nel migliorare la precisione del modello, le selezioni delle regole di combinazione vengono ignorate. Il boosting utilizza sempre un voto di maggioranza pesata per calcolare il punteggio degli obiettivi continui categoriali e una mediana pesata per calcolare il punteggio degli obiettivi continui.

Boosting e bagging. Specifica il numero di modelli di base da creare quando un obiettivo consiste nel migliorare la precisione o la stabilità del modello; per il bagging, si tratta del numero di campioni di bootstrap. Deve essere un numero intero positivo.

Avanzate

Replica risultati. L'impostazione di un seme casuale consente di replicare le analisi. Il generatore di numeri casuali viene utilizzato per scegliere i record presenti nell'insieme di prevenzione del sovradattamento. Specificare un numero intero o fare clic su **Genera** per creare un numero intero pseudo-casuale compreso tra 1 e 2147483647, compresi. Il valore predefinito è 54752075.

Opzioni modello

Salva valori previsti nel dataset. Il nome predefinito della variabile è *PredictedValue*.

Esporta modello. Questo scrive il modello in un file `.zip` esterno. È possibile utilizzare questo file modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio. Specificare un nome file valido e univoco. Se la specifica del file si riferisce ad un file esistente, il file viene sovrascritto.

Riepilogo del modello

La visualizzazione Riepilogo del modello è un'istantanea, un riepilogo immediato del modello.

Tabella. La tabella identifica alcune impostazioni del modello di alto livello, tra cui:

- Il nome della destinazione specificata nella scheda Campi,
- Se è stata eseguita la preparazione automatica dei dati come specificato nelle impostazioni Di base,
- Il metodo di selezione del modello e il criterio di selezione specificati nelle impostazioni Selezione modello. Inoltre viene visualizzato il valore del criterio di selezione per il modello finale, che viene presentato in un formato che predilige le dimensioni più piccole.

Grafico. Il grafico visualizza la precisione del modello finale, che viene presentato in un formato che predilige le dimensioni più grandi. Il valore è $100 \times R^2$ adattato per il modello finale.

Preparazione automatica dati

Questa vista mostra le informazioni su quali campi sono esclusi e su come vengono derivati i campi trasformati nel passo ADP (automatic data preparation, preparazione automatica dati). Per ogni campo che è stato trasformato o escluso, la tabella elenca il nome del campo, il suo ruolo nell'analisi e l'azione eseguita dal passo ADP. I campi sono disposti in ordine alfabetico crescente dei nomi dei campi. Le azioni che è possibile eseguire su ciascun campo includono:

- **Deriva durata: mesi** calcola il tempo trascorso, in mesi, dai valori in un campo contenente le date fino alla data corrente del sistema.
- **Deriva durata: ore** calcola il tempo trascorso, in ore, dai valori in un campo contenente le ore fino alla data corrente del sistema.
- **Modifica livello di misurazione da continuo a ordinale** ricompone i campi continui con meno di 5 valori univoci come campi ordinali.
- **Modifica livello di misurazione da ordinale a continuo** ricompone i campi ordinali con più di 10 valori univoci come campi continui.
- **Ritaglia valori anomali** imposta i valori dei predittori continui che si trovano oltre un valore di interruzione (3 deviazioni standard dalla media) per il valore di interruzione.

- **Sostituisci valori mancanti** sostituisce i valori mancanti dei campi nominali con la moda, dei campi ordinali con la mediana e dei campi continui con la media.
- **Unisci categorie per aumentare al massimo l'associazione all'obiettivo** identifica le categorie di predittori "simili" in base alla relazione tra l'input e l'obiettivo. Le categorie che non sono significativamente differenti (ovvero, che hanno un valore *p* maggiore di 0.05) vengono unite.
- **Escludi predittore costante / dopo gestione dei valori anomali / dopo l'unione di categorie** rimuove i predittori che hanno un solo valore, probabilmente dopo che sono state intraprese altre azioni ADP.

Importanza predittore

Di solito è opportuno concentrare la modellazione sui campi predittori più rilevanti, lasciando perdere o ignorando i meno importanti. In questo senso può essere utile il grafico dell'importanza dei predittori, che indica l'importanza relativa di ciascun predittore nella stima del modello. Dal momento che i valori sono relativi, la somma dei valori di tutti i predittori visualizzati è pari a 1,0. L'importanza dei predittori non ha nulla a che vedere con la precisione del modello. Riguarda unicamente l'importanza di ciascun predittore per l'elaborazione di una previsione, non il grado di precisione di quest'ultima.

Previsioni e osservazioni

Visualizza un grafico a dispersione raccolto dei valori previsti sull'asse verticale in base ai valori osservati sull'asse orizzontale. Idealmente, i punti devono giacere su una linea di 45 gradi; questa vista può indicare eventuali record previsti in modo particolarmente non corretto dal modello.

Residui

Visualizza un grafico diagnostico dei residui del modello.

Stili del grafico. Sono disponibili diversi stili di visualizzazione, che sono accessibili dall'elenco a discesa **Stile**.

- **Istogramma.** Si tratta di un istogramma raccolto dei residui studentizzati con una sovrapposizione della distribuzione normale. I modelli lineari presumono che i residui abbiano una distribuzione normale, quindi l'istogramma idealmente deve avvicinarsi molto alla linea uniforme (piatta).
- **Grafico P-P.** Si tratta di un grafico probabilità-probabilità raccolto che confronta i residui studentizzati con una distribuzione normale. Se la pendenza dei punti tracciati è meno ripida della linea normale, i residui mostrano più variabilità di una distribuzione normale; se la pendenza è più ripida, i residui mostrano meno variabilità di una distribuzione normale. Se i punti tracciati hanno una curva a forma di S, la distribuzione dei residui è asimmetrica.

Valori anomali

Questa tabella elenca i record che esercitano un'influenza non necessaria sul modello e visualizza l'ID record (se specificato nella scheda Campi), il valore obiettivo e la distanza di Cook. La distanza di Cook è una misura di quando i residui di tutti i record cambiano se un determinato record è stato escluso dal calcolo dei coefficienti del modello. Una distanza di Cook di grandi dimensioni indica che l'esclusione di un record dal calcolo cambia notevolmente i coefficienti, pertanto deve essere considerata influente.

I record influenti devono essere esaminati con attenzione per determinare se è possibile assegnare ad essi meno peso nella stima del modello o troncare i valori anomali ad una soglia accettabile oppure rimuovere completamente i record influenti.

Effetti

Questa vista mostra la dimensione di ciascun effetto nel modello.

Stili. Sono disponibili diversi stili di visualizzazione, che sono accessibili dall'elenco a discesa **Stile**.

- **Diagramma.** Si tratta di un grafico in cui gli effetti vengono ordinati dall'alto verso il basso, riducendo l'importanza del predittore. Le linee di connessione nel diagramma sono pesate in base alla significatività dell'effetto, per cui la larghezza della linea più grande corrisponde agli effetti più

significativi (valori *p* più piccoli). Passando con il cursore del mouse su una linea di connessione viene visualizzato un suggerimento che mostra il valore *p* e l'importanza dell'effetto. È l'impostazione predefinita.

- **Tabella.** Si tratta di una tabella ANOVA per gli effetti del modello globale e del modello individuale. Gli effetti individuali vengono ordinati dall'alto verso il basso, riducendo l'importanza del predittore. Si osservi che per impostazione predefinita la tabella è compressa per mostrare solo i risultati relativi al modello globale. Per visualizzare i risultati relativi agli effetti del modello individuale, fare clic sulla cella **Modello corretto** nella tabella.

Importanza predittore. Esiste un dispositivo di scorrimento Importanza predittore che controlla quali predittori vengono visualizzati nella vista. Questo non cambia il modello, ma consente semplicemente di concentrare l'attenzione sui predittori più importanti. Per impostazione predefinita, vengono visualizzati i primi 10 effetti.

Significatività. Esiste un dispositivo di scorrimento Significatività che controlla ulteriormente quali effetti vengono mostrati nella vista oltre a quelli mostrati in base all'importanza del predittore. Gli effetti con valori di significatività superiori al valore del dispositivo di scorrimento vengono nascosti. Questo non cambia il modello, ma consente semplicemente di concentrare l'attenzione sugli effetti più importanti. Per impostazione predefinita, il valore è 1.00, quindi non viene filtrato alcun effetto in base alla significatività.

Coefficienti

Questa vista mostra il valore di ciascun coefficiente nel modello. Si osservi che i fattori (predittori categoriali) sono indicatori codificati nel modello, in tal modo gli **effetti** contenenti i fattori avranno, in genere, più **coefficienti** associati; uno per ogni categoria, ad eccezione della categoria corrispondente al parametro ridondante (riferimento).

Stili. Sono disponibili diversi stili di visualizzazione, che sono accessibili dall'elenco a discesa **Stile**.

- **Diagramma.** Si tratta di un grafico che visualizza prima l'intercetta e poi ordina gli effetti dall'alto verso il basso, riducendo l'importanza del predittore. All'interno degli effetti che contengono fattori, i coefficienti sono disposti in ordine crescente di valori dei dati. Le linee di connessione nel diagramma sono colorate in base al segno del coefficiente (vedere la chiave del diagramma) e pesate in base alla significatività del coefficiente; la larghezza della linea più grande corrisponde ai coefficienti più significativi (valori *p* più piccoli). Passando con il cursore del mouse su una linea di connessione viene visualizzato un suggerimento che mostra il valore del coefficiente, il suo valore *p* e l'importanza dell'effetto a cui è associato il parametro. Questo è lo stile predefinito.
- **Tabella.** Mostra i valori, i test di significatività e gli intervalli di confidenza per i singoli coefficienti del modello. Dopo l'intercetta, gli effetti vengono ordinati dall'alto verso il basso, riducendo l'importanza del predittore. All'interno degli effetti che contengono fattori, i coefficienti sono disposti in ordine crescente di valori dei dati. Si osservi che per impostazione predefinita la tabella è compressa per mostrare solo il coefficiente, la significatività e l'importanza di ciascun parametro del modello. Per visualizzare l'errore standard, la statistica *t* e l'intervallo di confidenza, fare clic sulla cella **Coefficiente** nella tabella. Passando con il cursore del mouse sul nome di un parametro del modello nella tabella, viene visualizzato un suggerimento che mostra il nome del parametro, l'effetto a cui è associato il parametro e (per i predittori categoriali) le etichette dei valori associate al parametro del modello. Ciò può essere particolarmente utile per visualizzare le nuove categorie create quando la preparazione automatica dei dati unisce categorie simili di un predittore categoriale.

Importanza predittore. Esiste un dispositivo di scorrimento Importanza predittore che controlla quali predittori vengono visualizzati nella vista. Questo non cambia il modello, ma consente semplicemente di concentrare l'attenzione sui predittori più importanti. Per impostazione predefinita, vengono visualizzati i primi 10 effetti.

Significatività. Esiste un dispositivo di scorrimento Significatività che controlla ulteriormente quali coefficienti vengono mostrati nella vista oltre a quelli mostrati in base all'importanza del predittore. I coefficienti con valori di significatività superiori al valore del dispositivo di scorrimento vengono nascosti. Questo non cambia il modello, ma consente semplicemente di concentrare l'attenzione sui coefficienti più importanti. Per impostazione predefinita, il valore è 1.00, quindi non viene filtrato alcun coefficiente in base alla significatività.

Medie stimate

Si tratta dei grafici visualizzati per i predittori significativi. Il grafico visualizza il valore stimato dal modello di obiettivo sull'asse verticale per ciascun valore del predittore sull'asse orizzontale, mantenendo costanti tutti gli altri predittori. Offre una visualizzazione utile degli effetti dei coefficienti di ciascun predittore sull'obiettivo.

Nota: se nessuno dei predittori è significativo, non viene generata alcuna media stimata.

Riepilogo creazione modello

Quando si sceglie un algoritmo di selezione del modello diverso da **Nessuno** nelle impostazioni di Selezione modello, vengono forniti alcuni dettagli del processo di creazione del modello.

Stepwise in avanti. Quando Stepwise in avanti è l'algoritmo di selezione, la tabella visualizza le ultime 10 fasi nell'algoritmo Stepwise in avanti. Per ogni fase vengono mostrati il valore del criterio di selezione e gli effetti nel modello nella fase in questione. In questo modo, si ha un'idea del contributo dato da ciascuna fase nel modello. Ogni colonna consente di ordinare le righe, in modo che sia più facile vedere quali effetti sono presenti nel modello in una determinata fase.

Sottoinsiemi migliori. Quando Sottoinsiemi migliori è l'algoritmo di selezione, la tabella visualizza i primi 10 modelli. Per ogni modello vengono mostrati il valore del criterio di selezione e gli effetti nel modello. In tal modo, si ha un'idea della stabilità dei primi modelli; se tendono ad avere molti effetti simili con poche differenze, si può essere abbastanza fiduciosi sul "primo" modello; se tendono ad avere effetti molto diversi, alcuni degli effetti potrebbero essere troppo simili e devono essere combinati (o uno di essi rimosso). Ogni colonna consente di ordinare le righe, in modo che sia più facile vedere quali effetti sono presenti nel modello in una determinata fase.

Regressione lineare

La regressione lineare consente di stimare i coefficienti dell'equazione lineare, includendo una o più variabili indipendenti, che prevedono al meglio il valore della variabile dipendente. Ad esempio, è possibile tentare di prevedere le vendite annuali di un rappresentante (la variabile dipendente) in base a variabili indipendenti quali l'età, gli studi e gli anni di esperienza lavorativa.

Esempio. Il numero di partite vinte da una squadra di basket in una stagione è correlato al numero medio di punti effettuati dalla squadra per partita? Un grafico a dispersione indica che queste variabili sono correlate in modo lineare. Il numero di partite vinte e il numero medio di punti effettuati dalla squadra avversaria sono anch'essi correlati in modo lineare. Queste variabili hanno una relazione negativa. Al crescere del numero delle partite vinte, diminuisce il numero medio di punti effettuati dall'avversario. Con la regressione lineare, è possibile modellare la relazione di queste variabili. È possibile utilizzare un modello valido per stimare quante partite vinceranno le squadre.

Statistiche. Per ogni variabile: numero di casi validi, media e deviazione standard. Per ogni modello: coefficienti di regressione, matrice di correlazione, correlazioni di ordine zero e parziali, *R* multiplo, R^2, R^2 adattato, modifica in R^2, errore standard della stima, tabella di analisi della varianza, valori previsti e residui. Inoltre, intervalli di confidenza del 95% per ogni coefficiente di regressione, matrice della varianza-covarianza, fattore di inflazione della varianza, tolleranza, test di Durbin-Watson, misure della distanza (Mahalanobis, Cook e valori di leva), DfBeta, DfFit, intervalli di previsione e informazioni di diagnostica casewise. Grafici: grafici a dispersione, grafici parziali, istogrammi e grafici di probabilità normale.

Considerazioni sui dati della regressione lineare

Dati. Le variabili dipententi e indipendenti devono essere quantitative. È necessario che le variabili categoriali, come la religione, l'età o la regione di residenza, siano ricodificate come variabili binarie (dummy) o altri tipi di variabili di contrasto.

Ipotesi. Per ciascun valore della variabile indipendente, la distribuzione della variabile dipendente deve essere normale. La varianza della distribuzione della variabile dipendente deve essere costante per tutti i valori della variabile indipendente. La relazione tra la variabile dipendente e ogni variabile indipendente deve essere lineare e tutte le osservazioni devono essere indipendenti.

Per ottenere un'analisi della regressione lineare

1. Dai menu, scegliere:

   Analizza > **Regressione** > **Lineare...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Nella finestra di dialogo Regressione lineare, selezionare una variabile dipendente numerica.
3. Selezionare una o più variabili indipendenti numeriche.

Facoltativamente, è possibile:

- Raggruppare le variabili indipendenti in blocchi e specificare metodi di inserimento differenti per sottogruppi di variabili diversi.
- Scegliere una variabile di selezione per limitare l'analisi a un sottoinsieme di casi con valori particolari per questa variabile.
- Selezionare una variabile di casi per l'identificazione di punti nei grafici.
- Selezionare una variabile peso per WLS per un'analisi dei minimi quadrati pesati.

WLS. Consente di ottenere un modello di minimi quadrati ponderati. I punti dati vengono pesati in base al reciproco della loro varianza. Le osservazioni con varianza elevata hanno un peso minore nell'analisi rispetto a quelle con varianza ridotta. Se il valore della variabile peso è zero, negativo o mancante, il caso viene escluso dall'analisi.

Metodi di selezione della variabile di regressione lineare

La selezione del metodo consente di specificare come vengono inserite nell'analisi le variabili indipendenti. Utilizzando diversi metodi, è possibile creare molteplici modelli di regressione dallo stesso insieme di variabili.

- *Enter (Regressione)*. Una procedura per la selezione delle variabili nella quale tutte le variabili di un blocco sono inserite in un unico passo.
- *Stepwise*. Ad ogni passo, viene inserita la variabile indipendente non presente nell'equazione che ha la più piccola probabilità di F, se tale probabilità è sufficientemente piccola. Le variabili già presenti nell'equazione di regressione vengono rimosse se la loro probabilità di F diviene sufficientemente elevata. Il metodo termina quando nessun'altra variabile rispetta il criterio di inserimento o quello di rimozione.
- *Rimuovi*. Una procedura per la selezione delle variabili nella quale tutte le variabili di un blocco sono rimosse in un unico passo.
- *Eliminazione all'indietro*. Una procedura di selezione delle variabili in cui tutte le variabili vengono inserite nell'equazione e quindi rimosse in modo sequenziale. La variabile con la più bassa correlazione parziale rispetto alla variabile dipendente viene considerata la prima da rimuovere e viene rimossa se soddisfa il criterio di eliminazione. Dopo la rimozione della prima variabile, la variabile con la più bassa correlazione parziale tra quelle rimaste nell'equazione viene considerata come la prossima da eliminare. La procedura termina quando nell'equazione nessuna variabile soddisfa il criterio di rimozione.
- *Selezione in avanti*. Una procedura di selezione delle variabili a fasi in cui le variabili vengono inserite in modo sequenziale nel modello. La prima variabile da inserire nell'equazione è quella con la più elevata correlazione positiva o negativa con la variabile dipendente. Questa variabile viene inserita nell'equazione solo se soddisfa il criterio di inserimento. Se è stata inserita la prima variabile, viene considerata come successiva la variabile indipendente non presente nell'equazione che ha la più elevata correlazione parziale. La procedura termina quando non ci sono più variabili che soddisfano il criterio di inserimento.

I valori di significatività dell'output si basano sull'adattamento di un singolo modello. Pertanto, i valori di significatività in genere non sono validi quando viene utilizzato un metodo stepwise (stepwise, avanti o indietro).

Tutte le variabili devono soddisfare il criterio di tolleranza per essere inserite nell'equazione, indipendentemente dal metodo di inserimento specificato. Il livello di tolleranza predefinito è 0,0001. Una variabile non viene inserita se può far sì che la tolleranza di un'altra variabile già nel modello non rientri nel criterio di tolleranza già stabilito.

Tutte le variabili indipendenti selezionate vengono aggiunte a un solo modello di regressione. È tuttavia possibile specificare diversi metodi di inserimento per diversi sottoinsiemi di variabili. Ad esempio, è possibile inserire un blocco di variabili nel modello di regressione utilizzando la selezione stepwise e un secondo blocco utilizzando la selezione in avanti. Per aggiungere un secondo blocco di variabili a un modello di regressione, fare clic su **Avanti**.

Regressione lineare: Imposta regola

Nell'analisi verranno inseriti i casi definiti dalla regola di selezione impostata. Ad esempio, se si seleziona la variabile **Uguale a** e si immette 5 per il valore, nell'analisi verranno inclusi solo i casi in cui la variabile selezionata ha un valore uguale a 5. È inoltre possibile specificare un valore stringa.

Regressione lineare: grafici

I grafici possono facilitare la convalida delle ipotesi di normalità, linearità e uguaglianza delle varianze. I grafici sono inoltre utili per la rilevazione di valori anomali, osservazioni insolite e casi di influenza. Dopo averli salvati come nuove variabili, nell'Editor dei dati sono disponibili i valori previsti, i residui e altre informazioni di diagnostica per creare dei grafici con le variabili indipendenti. Sono disponibili i seguenti grafici:

Grafici a dispersione. È possibile tracciare due qualsiasi dei seguenti grafici: variabile dipendente, valori previsti standardizzati, residui standardizzati, residui eliminati, valori previsti adattati, residui studentizzati o residui eliminati studentizzati. Rappresentare nel grafico i residui standardizzati e i valori attesi standardizzati per verificare la linearità e l'uguaglianza delle varianze.

Elenco di variabili origine.Elenca la variabile dipendente (DEPENDNT) e le seguenti variabili (valori stimati e residui): valori stimati standardizzati (*ZPRED), residui standardizzati (*ZRESID), residui cancellati (*DRESID), valori stimati corretti (*ADJPRED), residui studentizzati (*SRESID), residui cancellati studentizzati (*SDRESID).

Produci tutti i grafici parziali. Consente di visualizzare i grafici a dispersione dei residui di ogni variabile indipendente e i residui della variabile dipendente quando entrambe le variabili sono regresse separatamente dal resto delle variabili indipendenti. Per la creazione di un grafico parziale è necessario che nell'equazione siano rappresentate almeno due variabili indipendenti.

Grafici dei residui standardizzati . È possibile ottenere gli istogrammi dei residui standardizzati e i grafici di probabilità normale confrontando la distribuzione dei residui standardizzati con una distribuzione normale.

Se vengono richiesti grafici, verranno visualizzate statistiche di riepilogo per i valori attesi standardizzati e per i residui standardizzati (***ZPRED** e ***ZRESID**).

Regressione lineare: salvataggio di nuove variabili

È possibile salvare i valori previsti, i residui e altre statistiche utili per le informazioni di diagnostica. Ogni selezione aggiunge una o più nuove variabili al file di dati attivo.

Valori previsti. I valori previsti dal modello di regressione per ogni caso.

- *Non standardizzati*.Il valore che il modello prevede per la variabile dipendente.
- *Standardizzati*.Una trasformazione di ogni valore previsto nella sua forma standardizzata. Ossia, il valore previsto medio viene sottratto dal valore previsto e la differenza viene divisa per la deviazione standard dei valori previsti. I valori previsti standardizzati hanno una media pari a 0 e una deviazione standard pari a 1.
- *Adattati.*.Il valore atteso per un caso quando quel caso è escluso dal calcolo dei coefficienti di regressione.

- *Errore standard delle previsioni delle medie*.Errori standard dei valori previsti. Una stima della deviazione standard del valore medio della variabile dipendente per i casi che hanno gli stessi valori delle variabili indipendenti.

Distanze. Misure per l'identificazione dei casi con combinazioni di valori insolite per le variabili indipendenti e dei casi che possono avere un notevole peso sul modello di regressione.

- *Di Mahalanobis*.Una misura di quanto i valori di un caso sulle variabili indipendenti differiscono dalla media di tutti i casi. Un'elevata distanza di Mahalanobis identifica un caso come caratterizzato da valori estremi per una o più variabili indipendenti.
- *Di Cook*.Una misura di quanto cambierebbero i residui di tutti i casi se un particolare caso fosse escluso dal calcolo dei coefficienti di regressione. Un valore alto del D di Cook indica che l'esclusione di un caso dal calcolo delle statistiche di regressione modifica sostanzialmente i coefficienti.
- *Valori di leva*.Misura l'influenza di un punto sull'adattamento della regressione. La leva centrata è compresa tra 0 (nessuna influenza sull'adattamento) e (N-1)/N.

Intervalli di previsione. I limiti superiore ed inferiore per gli intervalli di previsione singoli e medi.

- *Media*. Limiti inferiore e superiore (due variabili) per l'intervallo di previsione della risposta media prevista.
- *Singolo*.Limiti inferiore e superiore (due variabili) dell'intervallo di previsione della variabile dipendente per un unico caso.
- *Intervallo di confidenza*. Immettere un valore compreso tra 1 e 99.99 per specificare il livello di confidenza per i due intervalli di previsione. Per disporre di questa opzione è necessario aver selezionato Media o Individuale. I valori di intervallo di confidenza tipici sono 90, 95 e 99.

Residui. Il valore effettivo della variabile dipendente meno il valore atteso dall'equazione di regressione.

- *Non standardizzati*.La differenza tra un valore osservato e il valore stimato dal modello.
- *Standardizzati*.Il residuo diviso per una stima della sua deviazione standard. I residui standardizzati, noti anche come residui di Pearson, hanno una media pari a 0 e una deviazione standard pari a 1.
- *Studentizzati*.Il residuo diviso per una stima della sua deviazione standard che varia da caso a caso, a seconda della distanza dei valori di ogni caso sulle variabili indipendenti dai mezzi delle variabili indipendenti. A volte indicato come residui internamente studentizzati.
- *Eliminato*.Il residuo per un caso quando tale caso è escluso dal calcolo dei coefficienti di regressione. È la differenza tra il valore della variabile dipendente e il valore previsto adattato.
- *Per cancellazione studentizzati*.Il residuo eliminato per un caso diviso per il relativo errore standard. La differenza tra un residuo cancellato Studentizzato e il relativo residuo Studentized indica quanta differenza eliminerà un caso fa sulla propria previsione. A volte indicato come residui di studio esternamente.

Statistiche di influenza. La modifica nei coefficienti di regressione (DiffBeta) e nei valori attesi (DiffAdatt) che risultano dall'esclusione di un particolare caso. I valori DiffBeta e DiffAdatt standardizzati sono anche disponibili con il rapporto di covarianza.

- *DfBetas*. La differenza nel valore beta è la variazione del coefficiente di regressione risultante dall'esclusione di uno specifico caso. Un valore viene calcolato per ogni termine nel modello, compreso quello costante.
- *DfBeta standardizzata*. Differenza standardizzata nel valore beta. La variazione di un coefficiente di regressione quando un caso viene rimosso dall'analisi. Possono essere esaminati i casi con valore assoluto superiore a 2 diviso per la radice quadrata di N, dove N è il numero di casi. Un valore viene calcolato per ogni termine nel modello, compreso quello costante.
- *DfFit*. La differenza nel valore adattato è il cambiamento del valore previsto quando un caso viene escluso dall'analisi.
- *DiffAdatt standardizzata*. Differenza standardizzata nel valore di adattamento. La variazione del valore stimato quando un caso viene rimosso dall'analisi. Si possono esaminare valori standardizzati maggiori in valore assoluto a 2 per la radice quadrata di p/N, dove p è il numero di parametri del modello e N è il numero di casi.

- *Rapporto di covarianza.*.Il rapporto tra il determinante della matrice di covarianza con un caso particolare escluso dal calcolo dei coefficienti di regressione e il determinante della matrice di covarianza con tutti i casi inclusi. Se il rapporto è prossimo a 1, il caso non modifica in modo significativo la matrice di covarianza.

Statistiche dei coefficienti. Salva i coefficienti di regressione in un file di dati o datset. I dataset possono anche essere riutilizzati nella stessa sessione, ma non vengono salvati come file a meno che siano stati salvati come tali al termine della sessione. I nomi dei dataset devono essere conformi alle regole dei nomi delle variabili.

Esporta informazioni modello in file XML. Le stime del parametro e, se si desidera, le relative covarianze vengono esportati nel file specificato in formato XML (PMML). È possibile utilizzare questo file modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio.

Regressione lineare: Statistiche

Sono disponibili le seguenti statistiche:

Coefficienti di regressione - Stime visualizza il coefficiente di regressione *B*, errore standard di *B*, coefficiente standardizzato beta, valore *t* per *B*e livello di significatività a due code di *t*. **Intervalli di confidenza** visualizza gli intervalli di confidenza con il livello di confidenza specificato per ciascun coefficiente di regressione o una matrice di covarianza. **Matrice di covarianza** consente di visualizzare una matrice di varianza-covarianza dei coefficienti di regressione con le covarianze esterne alla diagonale e le varianze sulla diagonale. Viene inoltre visualizzata una matrice di correlazione.

Adattamento del modello - Le variabili immesse e rimosse dal modello sono elencate e vengono visualizzate le seguenti 'statistiche di bontà dell'adattamento': più *R*, R^2 e R^2rettificate, errore standard della stima e tabella di analisi della varianza.

Variazione di R quadrato - La variazione nella statistica R^2 prodotta aggiungendo o eliminando una variabile indipendente. Se la variazione di R^2 associata ad una variabile è elevata, ciò significa che la variabile è un valido stimolatore della variabile dipendente.

Descrittivo - Fornisce il numero di casi validi, la media e la deviazione standard per ogni variabile nell'analisi. Vengono inoltre visualizzati una matrice di correlazione con un livello di significatività a una coda e il numero di casi per ogni correlazione.

Correlazione parziale. La correlazione tra la variabile dipendente e una variabile indipendente quando gli effetti lineari delle altre variabili indipendenti nel modello sono stati rimossi dalla variabile indipendente. Correlata alla variazione di R-quadrato quando una variabile viene aggiunta a un'equazione. A volte detta correlazione semiparziale.

Correlazione parziale. La correlazione che rimane tra due variabili dopo aver rimosso la correlazione dovuta alla loro mutua associazione con le altre variabili. La correlazione fra la variabile dipendente e una variabile indipendente dopo aver rimosso da entrambe gli effetti lineari delle altre variabili indipendenti nel modello.

Diagnostica della collinearità - La collinearità (o multicollinearità) è la situazione indesiderata in cui una variabile indipendente è una funzione lineare di altre variabili indipendenti. Consente di visualizzare gli autovalori della matrice cross-product scalata e non centrata, l'indice di collinearità e le proporzioni della decomposizione della varianza con i fattori di inflazione della varianza (VIF) e le tolleranze per le singole variabili.

Criteri di selezione - Include il criterio di informazione di Akaike (AIC), il criterio di previsione di Ameniya (PC), l'errore quadratico medio condizionale di Mallows del criterio di previsione (Cp) e il criterio bayesiano di Schwarz (SBC). Le statistiche vengono visualizzate nella tabella Riepilogo modello.

Residui: è possibile selezionare **'Statistica PRESS'** da utilizzare come statistica di convalida incrociata per confrontare modelli diversi. Viene visualizzato anche il test **'Durban -Watson'** per la correlazione seriale dei residui. Scegliere le informazioni **'Diagnostica per casi'** per i casi che soddisfano il criterio di selezione (valori anomali superiori a *n* deviazioni standard).

Regressione lineare: Opzioni

Sono disponibili le seguenti opzioni:

Criteri di accettazione e rifiuto. Queste opzioni si applicano quando viene specificato il metodo di selezione delle variabili in avanti, all'indietro o a fasi. Le variabili possono essere inserite o rimosse dal modello in base alla significatività (probabilità) del valore *F* o al valore *F* stesso.

- *Usa probabilità di F*. Una variabile viene immessa nel modello se il livello di significatività del suo valore F è minore del valore di inserimento e viene rimossa se il livello di significatività è maggiore del valore di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi e Inserimento deve essere minore di Rimozione. Alzando il valore di inserimento e/o abbassando quello di rimozione Per rimuovere un numero maggiore di variabili dal modello, ridurre il valore di rimozione.
- *Utilizza valore F*. Una variabile viene immessa nel modello se il suo valore F è maggiore del valore di immissione e viene rimossa se il valore F è minore del valore di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi e Inserimento deve essere maggiore di Rimozione Abbassando il valore di inserimento e/o alzando quello di rimozione Per rimuovere un numero maggiore di variabili dal modello, aumentare il valore di rimozione.

Tolleranza. Per impostazione predefinita, il valore è .0001. La tolleranza è la proporzione della varianza di una variabile nell'equazione che non è spiegata da altre variabili indipendenti nell'equazione. La tolleranza minima per qualsiasi variabile nell'equazione se la variabile in esame è stata inclusa nell'analisi è la tolleranza minima di una variabile che non è inclusa nell'equazione. Le variabili devono superare i test di tolleranza e tolleranza minima per entrare e rimanere in un'equazione di regressione. Se una variabile supera i criteri di tolleranza, è idonea per l'inclusione in base al metodo in vigore.

Includi costante nell'equazione. Per impostazione predefinita, il modello di regressione include un termine costante. Se l'opzione è deselezionata, viene forzato il passaggio della curva di regressione per l'origine, il che avviene raramente. Alcuni risultati di una curva di regressione che passa per l'origine non sono confrontabili con i risultati della regressione che include una costante. Ad esempio R^2 non può essere interpretato nel modo usuale.

Valori mancanti. È possibile scegliere tra le opzioni seguenti:

- **Escludi casi listwise.** Sono inclusi nell'analisi solo i casi con valori validi per tutte le variabili.
- **Escludi casi pairwise.** Per calcolare il coefficiente di correlazione su cui si basa l'analisi della regressione vengono utilizzati i casi con dati completi per la coppia di variabili correlate. I gradi di libertà sono basati su *N* pairwise minimo.
- **Sostituisci con la media.** Per i calcoli vengono utilizzati tutti i casi e la media della variabile viene sostituita alle osservazioni mancanti.

Funzioni aggiuntive del comando REGRESSION

Il linguaggio della sintassi dei comandi consente inoltre di:

- Scrivere una matrice di correlazione o leggere una matrice anziché i dati grezzi per ottenere l'analisi di regressione (con il sottocomando `MATRIX`).
- Specificare i livelli di tolleranza (tramite il sottocomando `CRITERIA`).
- Ottenere più modelli per variabili dipendenti uguali o diverse (con i sottocomandi `METHOD` e `DEPENDENT`).
- Ottenere statistiche aggiuntive (con i sottocomandi `DESCRIPTIVES` e `STATISTICS`).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Regressione ordinale

La procedura Regressione ordinale consente di definire la dipendenza di una risposta ordinale politomica in un insieme di stimatori, che possono essere fattori o covariate. Il disegno della regressione ordinale è basato sulla metodologia di McCullagh (1980, 1998) e la procedura è denominata `PLUM` nella sintassi.

L'analisi della regressione lineare standard comporta la riduzione al minimo della somma delle differenze al quadrato tra una variabile di risposta (dipendente) e una combinazione pesata di variabili stimatore (indipendenti). I coefficienti stimati riflettono il modo in cui le modifiche dei predittori influiscono sulla risposta. Si presume che la risposta sia numerica e ciò significa che le modifiche al livello di risposta sono uguali nell'intervallo della risposta. Ad esempio, la differenza di altezza tra una persona alta 150 cm e una persona alta 140 cm è di 10 cm ed equivale alla differenza di altezza tra una persona alta 210 cm e una persona alta 200 cm. Queste relazioni non sono necessariamente valide per le variabili ordinali, nelle quali la scelta e il numero delle categorie di risposta possono essere arbitrarie.

Esempio. È possibile utilizzare la regressione ordinale per studiare la reazione dei pazienti a un dosaggio medicinale. Le possibili reazioni possono essere classificate come *nessuna*, *lieve*, *moderata* o *grave*. La differenza tra una reazione lieve e una moderata è molto difficile o impossibile da quantificare ed è basata sulla percezione. In generale, la differenza tra una risposta lieve e una moderata può essere maggiore o minore della differenza tra una risposta moderata e una grave.

Statistiche e grafici. Frequenze osservate e attese e frequenze cumulative, residui di Pearson per le frequenze e frequenze cumulative, probabilità osservate e attese, probabilità osservate e probabilità cumulative attese per ciascuna categoria di risposta in base al modello di covariata, correlazione asintotica e matrici di covarianza delle stime del parametro, chi-quadrato di Pearson e chi-quadrato del rapporto di verosimiglianza, statistiche della bontà di adattamento, cronologia delle iterazioni, test sull'ipotesi di linee parallele, stime del parametro, errori standard, intervalli di confidenza e statistiche R^2 di Cox e Snell, di Nagelkerke e di McFadden.

Considerazioni sui dati della procedura Regressione ordinale

Dati. Si presume che la variabile dipendente sia ordinale e può essere numerica o stringa. L'ordinamento è determinato dalla disposizione dei valori della variabile dipendente in ordine crescente, in cui il valore più basso definisce la prima categoria. Si presume che le variabili fattore siano categoriali, mentre le covariate devono essere numeriche. Si noti che l'utilizzo di più covariate continue in genere comporta la creazione di una tabella di probabilità di cella di dimensioni molto grandi.

Ipotesi. È consentita una sola variabile di risposta, che deve essere specificata. Inoltre, per ciascun modello distinto di valori nelle variabili indipendenti, si presume che le risposte siano variabili multinomiali indipendenti.

Procedure correlate. La regressione logistica nominale utilizza modelli simili per le variabili dipendenti nominali.

Come ottenere una regressione ordinale

1. Dai menu, scegliere:

   Analizza > **Regressione** > **Ordinale...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una variabile dipendente.
3. Fare clic su **OK**.

Regressione ordinale: Opzioni

Nella finestra di dialogo Opzioni è possibile adattare i parametri utilizzati nell'algoritmo di stima iterativo, scegliere un livello di confidenza per le stime del parametro e selezionare una funzione di collegamento.

Iterazioni. È possibile personalizzare l'algoritmo iterativo.

- **Numero massimo di iterazioni.** Specificare un intero non negativo. Se si specifica 0, la procedura restituisce le stime iniziali.
- **Max dimezzamenti.** Specificare un intero positivo.
- **Convergenza logaritmo della verosimiglianza.** L'algoritmo si interrompe se il cambiamento assoluto o relativo nel logaritmo della verosimiglianza è inferiore a questo valore. Il criterio non viene utilizzato se si specifica 0.

- **Convergenza parametri.** L'algoritmo si interrompe se il cambiamento assoluto o relativo in ciascuna delle stime del parametro è inferiore a questo valore. Il criterio non viene utilizzato se si specifica 0.

Intervallo di confidenza. Specificare un valore maggiore o uguale a 0 e minore di 100.

Delta. Valore aggiunto alle frequenze zero di cella. Specificare un valore non negativo inferiore a 1.

Tolleranza della singolarità. Utilizzata per controllare gli stimatori a dipendenza elevata. Selezionare un valore dall'elenco delle opzioni.

Funzione di collegamento. La funzione di collegamento è la trasformazione delle probabilità cumulative che permette di stimare il modello. Sono disponibili le cinque funzioni di collegamento riportate di seguito.

- **Logit.** $f(x) = \log(x/(1-x))$. Generalmente utilizzato per categorie distribuite in modo uniforme.
- **Log-log complementare.** $f(x)=\log(-\log(1-x))$. In genere, viene utilizzata quando le categorie più alte sono più probabili.
- **Log-log negativa.** $f(x)=-\log(-\log(x))$. In genere, viene utilizzata quando le categorie più basse sono più probabili.
- **Probit.** $f(x) = \Phi^{-1}(x)$. In genere, viene utilizzata quando la variabile latente viene distribuita normalmente.
- **Cauchit (Cauchy inverso).** $f(x)=\tan(\pi(x-0.5))$. In genere, viene utilizzata quando la variabile latente ha molti valori estremi.

Regressione ordinale: Output

Nella finestra di dialogo Output è possibile creare tabelle da visualizzare nel Visualizzatore e salvare variabili nel file di lavoro.

Visualizza. Consente di creare tabelle per:

- **Stampa cronologia delle iterazioni per ogni n fasi.** Stampa le stime del parametro e del logaritmo della verosimiglianza per la frequenza di iterazioni di stampa specificata. La prima e l'ultima iterazione vengono sempre stampate.
- **Statistiche della bontà di adattamento.** Statistica chi-quadrato di Pearson e dei rapporti di verosimiglianza. Vengono elaborate in base alla classificazione specificata nell'elenco di variabili.
- **Statistiche di riepilogo.** Statistiche R^2 di Cox e Snell, di Nagelkerke e di McFadden.
- **Stime dei parametri.** Stime del parametro, errori standard e intervalli di confidenza.
- **Correlazione asintotica delle stime del parametro.** Matrice delle correlazioni delle stime dei parametri.
- **Covarianza asintotica delle stime del parametro.** Matrice delle covarianze delle stime dei parametri.
- **Informazioni sulle celle.** Frequenze osservate e previste e frequenze cumulative, residui di Pearson per frequenze e frequenze cumulative, probabilità osservate e previste, probabilità cumulative osservate e previste per ciascuna categoria di risposta in base al modello di covariata. Si noti che per i modelli che includono più modelli di covariate, ad esempio i modelli con covariate continue, questa opzione può creare una tabella di dimensioni molto grandi e difficile da gestire.
- **Test di linee parallele.** Test dell'ipotesi di equivalenza dei parametri di ubicazione nei livelli della variabile dipendente. È disponibile per il modello di sola ubicazione.

Variabili salvate. Salva le seguenti variabili nel file di lavoro:

- **Probabilità di risposta stimate.** Probabilità stimate dal modello per la classificazione di un modello di fattore o covariata nelle categorie di risposta. Il numero di probabilità corrisponde al numero di categorie di risposta.
- **Categoria prevista.** La categoria di risposta che ha la maggiore probabilità stimata per un modello di fattore o covariata.

- **Probabilità di categoria prevista.** Probabilità stimata di classificazione di un modello di fattore o covariata nella categoria prevista. La probabilità corrisponde inoltre al massimo di probabilità stimate del modello di fattore o covariata.
- **Probabilità di categoria reale.** Probabilità stimata di classificazione di un modello di fattore o covariata nella categoria reale.

Stampa logaritmo della verosimiglianza. Controlla la visualizzazione del logaritmo della verosimiglianza. L'**inclusione della costante multinomiale** consente di ottenere il valore completo della verosimiglianza. Per confrontare i risultati dei prodotti che non includono la costante, si può scegliere di escluderla.

Regressione ordinale: Ubicazione

Nella finestra di dialogo Ubicazione è possibile specificare il modello di ubicazione per l'analisi.

Specifica modello. Un modello di effetti principali include gli effetti principali di covariate e fattori, ma non gli effetti di interazione. È possibile creare un modello personalizzato per specificare i sottoinsiemi di interazioni dei fattori o di interazioni di covariate.

Fattori/covariate. Vengono elencati i fattori e le covariate.

Modello ubicazione. Il modello dipende dagli effetti principali e dagli effetti di interazione selezionati.

Per i fattori e le covariate selezionati:

Interazione
Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

Effetti principali
Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti - 2 vie
Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti - 3 vie
Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti - 4 vie
Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti - 5 vie
Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Creazione di termini e termini personalizzati

Crea termini
Utilizzare questa opzione quando si desidera includere termini non nidificati di un certo tipo (ad esempio, effetti principali) per tutte le combinazioni di una serie selezionata di fattori e covariate.

Crea termini personalizzati
Utilizzare questa opzione quando si desidera includere termini nidificati o creare esplicitamente una variabile termine dalla variabile. La creazione di un termine nidificato richiede le seguenti operazioni:

Regressione ordinale: Scala

Nella finestra di dialogo Scala è possibile specificare il modello di scala per l'analisi.

Fattori/covariate. Vengono elencati i fattori e le covariate.

Modello scala. Il modello dipende dagli effetti di interazione e principali selezionati.

Per i fattori e le covariate selezionati:

Interazione
Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

Effetti principali
Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti - 2 vie
Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti - 3 vie
Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti - 4 vie
Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti - 5 vie
Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Creazione di termini e termini personalizzati

Crea termini
Utilizzare questa opzione quando si desidera includere termini non nidificati di un certo tipo (ad esempio, effetti principali) per tutte le combinazioni di una serie selezionata di fattori e covariate.

Crea termini personalizzati
Utilizzare questa opzione quando si desidera includere termini nidificati o creare esplicitamente una variabile termine dalla variabile. La creazione di un termine nidificato richiede le seguenti operazioni:

Funzioni aggiuntive del comando PLUM

Per personalizzare la procedura Regressione ordinale è possibile incollare le impostazioni selezionate in una finestra della sintassi e quindi modificare la sintassi del comando PLUM così ottenuta. Il linguaggio della sintassi dei comandi consente inoltre di:

- Creare test sull'ipotesi personalizzati specificando ipotesi nulla come combinazioni lineari dei parametri.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Regressione Elastic Net lineare

Linear Elastic Net utilizza la classe Python sklearn.linear_model.ElasticNet per stimare i modelli di regressione lineare regolarizzati per una variabile dipendente su una o più variabili indipendenti. La regolarizzazione combina le penalità L1 (Lasso) e L2 (Ridge). L'estensione include modalità facoltative per visualizzare i grafici di traccia per i diversi valori di alfa per un determinato rapporto L1 e per selezionare i valori del rapporto L1 e dell'iperparametro alfa in base alla convalida incrociata. Quando un modello singolo viene montato o crossvalidation viene utilizzato per selezionare il rapporto di penalità e / o alfa, una partizione di dati holdout può essere utilizzata per stimare le prestazioni fuori campione.

Oltre ad allestimento di un modello con valori specificati del rapporto di penalità L1 e parametro di regolarizzazione alfa, la rete elastica lineare può visualizzare un tracciato di tracciamento dei valori di coefficiente per una gamma di valori alfa per un determinato rapporto, oppure facilitare la scelta del valore dell'iperparametri tramite crossvalidation di k - fold su specifiche griglie di valori. Se viene adattato un singolo modello o viene eseguita la selezione del rapporto e / o alfa tramite convalida incrociata, il modello finale può essere applicato ai dati trattenuti creati da una partizione dei dati di input per ottenere una stima valida delle prestazioni fuori campione del modello.

Ottenimento di un'analisi Linear Elastic Net Regression

1. Dai menu, scegliere:

   Analizzare > **Regressione** > **Linear OLS Alternative** > **Elastic Net**

   La finestra di dialogo consente di specificare una variabile che assegna ogni caso nel dataset attivo al campione di formazione o holdout.

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una variabile di destinazione numerica. È necessaria una sola variabile di destinazione per eseguire un'analisi.
3. Specificare un dipendente numerico.
4. Specificare almeno una variabile di covariata variabile o numerica.

Facoltativamente, **Partizione** fornisce un modo per creare un sottoinsieme di holdout o di test dei dati di input per la stima delle prestazioni fuori campione del modello specificato o scelto. Tutto il partizionamento viene eseguito dopo la cancellazione listica di eventuali casi con dati non validi per qualsiasi variabile utilizzata dalla procedura. Si noti che per il crossvalidation, le pieghe o le partizioni dei dati formativi vengono create in Python. I dati di holdout creati dalla partizione non vengono utilizzati nella stima, indipendentemente dalla modalità in vigore.

La partizione può essere definita specificando il rapporto dei casi assegnati casualmente a ciascun campione (in **Partizioni di addestramento e di holdout**) oppure mediante una variabile che assegna ciascun caso al campione di addestramento o di holdout. Non è possibile specificare sia la formazione che le variabili. Se la partizione non viene specificata, viene creato un campione holdout di circa il 30% dei dati di input.

Il **Formazione%** specifica il numero relativo di casi nel dataset attivo per assegnare casualmente al campione di formazione. La formazione predefinita è di 70%.

Linear Elastic Net Regression: Opzioni

La scheda Opzioni fornisce opzioni per:

Modalità

Questa selezione fornisce opzioni per specificare una delle seguenti modalità:

Adatta a rapporto L1 specificato e alfa
Quando si seleziona questa modalità, un singolo modello viene adattato ai dati di addestramento utilizzando il rapporto L1 specificato e i valori di regolarizzazione alfa. È l'impostazione predefinita. Se viene specificata una partizione, il modello singolo o finale adattato viene applicato ai dati di test tenuti fuori per stimare le prestazioni fuori campione.

In **Grafico**, è possibile selezionare i grafici dei valori osservati e / o residui rispetto ai valori previsti.

Sotto **Salva**è possibile specificare valori e residui previsti da salvare.

Grafico delle tracce
Quando si seleziona questa modalità, tre grafici per i dati di addestramento vengono visualizzati come una funzione di alfa per l'insieme specificato di valori alfa:

- Una traccia di coefficienti di regressione.
- Un grafico di R^2.
- Una trama di errore quadrato medio (MSE).

Sebbene la partizione sia onorata, non vengono forniti risultati per i dati di test detenuti perché nessun modello finale risulta da questa modalità.

Rapporto L1 e / o selezione alfa tramite convalida incrociata
Quando si seleziona questa modalità, viene effettuata una ricerca a griglia con convalida incrociata per valutare i modelli e vengono scelti il miglior rapporto e i migliori valori alfa in base alla migliore media R^2 rispetto alle occorrenze di convalida. Il campo **Numero di occorrenze di convalida incrociata** può essere utilizzato per modificare il valore predefinito di cinque suddivisioni o occorrenze per la convalida incrociata. Se viene specificata una partizione, il modello singolo o finale adattato viene applicato ai dati di test tenuti fuori per stimare le prestazioni fuori campione.

In **Visualizzazione**, è possibile selezionare la visualizzazione delle informazioni di base solo sul modello con il valore scelto di rapporto e alfa (**Migliore**), delle informazioni di base su tutti i modelli confrontati (**Confronta modelli**) o delle informazioni complete su tutte le suddivisioni o le occorrenze per tutti i modelli (**Confronta modelli e occorrenze**). **Best** è il default.

In **Grafico**, è possibile selezionare i grafici della media R^2 e / o MSE sulle occorrenze di convalida. È possibile selezionare anche i grafici dei residui osservati e / o rispetto ai valori previsti.

Sotto **Salva**è possibile specificare valori e residui previsti da salvare.

Specifica singoli rapporti L1
Quando si seleziona questa modalit ... per la modalit ... **Adatta con il rapporto L1 specificato e la modalit ... alfa** o **Traccia del grafico** , Š possibile specificare un singolo valore del rapporto di penalit ... L1 . Quando si seleziona il rapporto **L1 e / o la selezione alfa mediante la modalità di convalida incrociata** , è possibile specificare più valori.

Specifica griglia di valori alfa
Quando si seleziona questa modalità per il rapporto **L1 e / o la selezione alfa mediante la modalità di convalida incrociata** , è possibile specificare una griglia di valori alfa univoci da un valore **Inizio** (`value1`) a un valore **Fine** (`value2`) con l'incremento di **Per** (`value3`). Se specificato, è consentita una sola serie valida di [`value1 TO value2 BY value3`]. Deve soddisfare tale `0 ≤ value1 ≤ value2 ≤ 1`. Nei casi in cui `value1 = value2`, equivale a specificare un singolo `value1`, indipendentemente da `value3`.

I grafici vengono visualizzati utilizzando la metrica specificata per gli assi X orizzontali dei valori alfa variabili.

Specifica singoli valori alfa
Quando si seleziona questa modalità per la modalità **Adatta con il rapporto L1 specificato e la modalità alfa** , è possibile specificare un singolo valore di regolarizzazione alfa. Se selezionato per la modalità **Traccia** o **L1 rapporto e / o selezione alfa tramite convalida incrociata** , è possibile specificare più valori.

Specifica griglia di valori alfa
Quando si seleziona questa modalità per la modalità **Traccia grafico** o **L1 rapporto e / o selezione alfa tramite convalida incrociata** , è possibile specificare una griglia di valori alfa univoci da un valore **Inizio** (`value1`) a un valore **Fine** (`value2`) con l'incremento di **Per** (`value3`). Se specificato, è consentita una sola serie valida di [`value1 TO value2 BY value3`]. Deve soddisfare tale `0 ≤ value1 ≤ value2 ≤ 1`. Nei casi in cui `value1 = value2`, equivale a specificare un singolo `value1`, indipendentemente da `value3`.

La **metrica Alpha** per intervalli di valori può essere **Linear** o **Base 10 logaritmica** (10 viene innalzato alla potenza dei valori specificati).

I grafici vengono visualizzati utilizzando la metrica specificata per gli assi X orizzontali dei valori alfa variabili.

Criteri

Analisi dei controlli.

Includi intercettazione
Questo criterio include un'intercettazione nei modelli adattati. Si noti che la procedura di estensione non centro o standardizza la variabile dipendente e l'intercettazione non è penalizzata durante la stima.

Standardizza predittori
Standardizza tutte le variabili indipendenti.

Numero di occorrenze di convalida incrociata
Il numero di suddivisioni o occorrenze per la valutazione della convalida incrociata dei modelli. Deve essere un valore intero positivo maggiore di 1. Il valore predefinito è 5.

Python stato casuale
Il valore dell'impostazione random_state in Python utilizzato durante la valutazione della convalida incrociata dei modelli. Consente la replica dei risultati che coinvolgono numeri pseudo - casuali. Il valore deve essere un numero intero compreso tra 0 e 2^{32}-1. Il valore predefinito è 0.

Limite di tempo (minuti)
Il numero di minuti consentiti per i calcoli del modello da eseguire. Se si specifica 0, il timer viene disattivato. Il valore predefinito è 5.

Visualizza

Questa opzione specifica la quantità di output da visualizzare per il rapporto **L1 e / o la selezione alfa tramite la modalità di convalida incrociata** .

Cordiali saluti
Visualizza solo i risultati di base per il modello migliore scelto. È impostato per impostazione predefinita.

Confronta modelli
Visualizza i risultati di base per tutti i modelli valutati.

Confronta modelli e occorrenze
Visualizza risultati completi verbose per ogni suddivisione o piega per ogni modello valutato.

Grafico

Questa opzione specifica i grafici dei valori osservati o residui rispetto ai valori previsti e, con la convalida incrociata, la specifica dei grafici dell'errore quadratico medio (MSE) e / o della media R^2 rispetto alle occorrenze di convalida incrociata rispetto ai valori alfa.

Media quadratica della convalida incrociata (MSE) rispetto a alfa
Per il rapporto **L1 e / o la selezione alfa tramite la modalità di convalida incrociata** , visualizza un grafico a linee della media MSE sulle occorrenze di convalida incrociata rispetto all'alfa per il valore del miglior rapporto L1 specificato o selezionato. Per la modalità **Traccia traccia** , una trama simile viene prodotta automaticamente in base ai dati di formazione completi.

Valore medio di R - quadrato con convalida incrociata rispetto a alfa
Per il rapporto **L1 e / o la selezione alfa mediante la modalità di convalida incrociata** , visualizza un grafico a linee della media R^2 sulle occorrenze di convalida incrociata rispetto all'alfa per il valore del miglior rapporto L1 specificato o selezionato. Per la modalità **Traccia traccia** , una trama simile viene prodotta automaticamente in base ai dati di formazione completi.

Osservato rispetto a Previsto
Visualizza un grafico a dispersione dei valori osservati rispetto a quelli previsti per il modello specificato o migliore.

Residui rispetto a Previsto
Visualizza un grafico a dispersione dei residui rispetto ai valori previsti per il modello specificato o migliore.

Salva

Specifica le variabili da salvare al dataset attivo.

Valori previsti
Salvare i valori previsti dal modello specificato o migliore al dataset attivo. È anche possibile specificare un **Nome variabile personalizzata**.

Residui
Salvare i residui dalle previsioni del modello specificate o migliori al dataset attivo. È anche possibile specificare un **Nome variabile personalizzata**.

Regressione Lasso lineare

Linear Lasso utilizza la classe Python sklearn.linear_model.Lasso per stimare i modelli di regressione lineare regolarizzata con perdita L1 per una variabile dipendente su una o più variabili indipendenti e include modalità facoltative per la visualizzazione dei grafici di traccia e per selezionare il valore dell'iperparametro alfa in base alla convalida incrociata. Quando viene montato un modello singolo o un crossvalidation viene utilizzato per selezionare alpha, una partizione di dati holdout può essere utilizzata per stimare le prestazioni fuori campione.

Oltre ad allestimento di un modello con un valore specificato del parametro di regolarizzazione alfa, il lazo lineare può visualizzare una traccia di tracciamento dei valori di coefficiente per una gamma di valori alfa, oppure facilitare la scelta del valore iperparametro tramite crossvalidation di k - fold su specifiche griglie di valori. Se viene adattato un singolo modello o viene eseguita la selezione alfa mediante convalida incrociata, il modello finale può essere applicato ai dati trattenuti creati da una partizione dei dati di input per ottenere una stima valida delle prestazioni fuori campione del modello.

Ottenimento di un'analisi Linear Lasso Regression

1. Dai menu, scegliere:

   Analizzare > **Regressione** > **Alternative OLS** > **Lazo**

   La finestra di dialogo consente di specificare una variabile che assegna ogni caso nel dataset attivo al campione di formazione o holdout.

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una variabile di destinazione numerica. È necessaria una sola variabile di destinazione per eseguire un'analisi.
3. Specificare un dipendente numerico.
4. Specificare almeno una variabile di covariata variabile o numerica.

Facoltativamente, **Partizione** fornisce un modo per creare un sottoinsieme di holdout o di test dei dati di input per la stima delle prestazioni fuori campione del modello specificato o scelto. Tutto il partizionamento viene eseguito dopo la cancellazione listica di eventuali casi con dati non validi per qualsiasi variabile utilizzata dalla procedura. Si noti che per il crossvalidation, le pieghe o le partizioni dei dati formativi vengono create in Python. I dati holdout creati dalla partizione non vengono utilizzati in stima, indipendentemente dalla modalità in vigore.

La partizione può essere definita sia specificando il rapporto dei casi assegnati casualmente ad ogni campione (sotto le partizioni **Formazione e Holdout**), sia da una variabile che assegna ogni caso al campione di formazione o holdout. Non è possibile specificare sia la formazione che le variabili. Se la partizione non viene specificata, viene creato un campione holdout di circa il 30% dei dati di input.

Il **Formazione%** specifica il numero relativo di casi nel dataset attivo per assegnare casualmente al campione di formazione. La formazione predefinita è di 70%.

Regressione Lineare Laica: Opzioni

La scheda Opzioni fornisce opzioni per:

Modalità

Questa selezione fornisce opzioni per specificare una delle seguenti modalità:

Adatta con alfa specificato
Quando si seleziona questa modalità, un singolo modello viene adattato ai dati di addestramento utilizzando un solo valore di regolarizzazione alfa. È impostato per impostazione predefinita. Se viene specificata una partizione, il modello singolo o finale adattato viene applicato ai dati di test tenuti fuori per stimare le prestazioni fuori campione.

In **Grafico**, è possibile selezionare i grafici dei valori osservati e / o residui rispetto ai valori previsti.

Sotto **Salva**è possibile specificare valori e residui previsti da salvare.

Grafico delle tracce
Quando si seleziona questa modalità, tre grafici per i dati di addestramento vengono visualizzati come una funzione di alfa per l'insieme specificato di valori alfa:

- Una traccia di coefficienti di regressione.
- Un grafico di R^2.
- Una trama di errore quadrato medio (MSE).

Sebbene la partizione sia onorata, non vengono forniti risultati per i dati di test detenuti perché nessun modello finale risulta da questa modalità.

Selezione alfa tramite convalida incrociata

Quando si seleziona una ricerca a griglia con convalida incrociata per valutare i modelli e si sceglie l'alfa migliore in base alla media migliore, R^2 rispetto alle occorrenze di convalida. Il campo **Numero di occorrenze di convalida incrociata** può essere utilizzato per modificare il valore predefinito di cinque suddivisioni o occorrenze per la convalida incrociata. Se viene specificata una partizione, il modello singolo o finale adattato viene applicato ai dati di test tenuti fuori per stimare le prestazioni fuori campione.

In **Visualizzazione**, è possibile selezionare la visualizzazione delle informazioni di base solo sul modello con il valore scelto di alfa (**Migliore**), delle informazioni di base su tutti i modelli confrontati (**Confronta modelli**) o delle informazioni complete su tutte le suddivisioni o le occorrenze per tutti i modelli (**Confronta modelli e occorrenze**). **Best** è il default.

In **Grafico**, è possibile selezionare i grafici della media R^2 e / o MSE sulle occorrenze di convalida. È possibile selezionare anche i grafici dei residui osservati e / o rispetto ai valori previsti.

Sotto **Salva**è possibile specificare valori e residui previsti da salvare.

Specifica singoli valori alfa
Quando si seleziona la modalità **Adatta con alfa specificato** , è possibile specificare un singolo valore di regolarizzazione alfa. Quando si seleziona la modalità **Traccia del grafico** o **Selezione alfa tramite convalida incrociata** , è possibile specificare più valori.

Valori
Specificare uno o più valori di regolarizzazione alfa positivi. I valori multipli possono essere specificati singolarmente o come intervalli. Il valore predefinito è 1.

Specifica griglia di valori alfa
Quando si seleziona la modalità **Traccia** o **Selezione alfa tramite convalida incrociata** , è possibile specificare una griglia di valori alfa univoci da un valore **Inizio** (`value1`) a un valore **Fine** (`value2`) con l'incremento di **Da** (`value3`). Se specificato, è consentita una sola serie valida di [`value1 TO value2 BY value3`]. Deve soddisfare tale $0 \leq value1 \leq value2 \leq 1$. Nei casi in cui $value1 = value2$, equivale a specificare un singolo `value1`, indipendentemente da `value3`.

La **metrica Alpha** per intervalli di valori può essere **Linear** o **Base 10 logaritmica** (10 viene innalzato alla potenza dei valori specificati).

I grafici vengono visualizzati utilizzando la metrica specificata per gli assi X orizzontali dei valori alfa variabili.

Criteri

Analisi dei controlli.

Includi intercettazione
Include un'intercettazione nel modello montato (s). Si noti che la procedura di estensione non centro o standardizza la variabile dipendente e l'intercettazione non è penalizzata durante la stima.

Standardizza predittori
Standardizza tutte le variabili indipendenti.

Numero di occorrenze di convalida incrociata
Il numero di suddivisioni o occorrenze per la valutazione della convalida incrociata dei modelli. Deve essere un valore intero positivo maggiore di 1. Il valore predefinito è 5.

Python stato casuale
Il valore dell'impostazione random_state in Python utilizzato quando si esegue la valutazione di convalida incrociata dei modelli. Consente la replica dei risultati che coinvolgono numeri pseudo - casuali. Deve essere un intero compreso nell'intervallo da 0 a 2^{32}-1. Il valore predefinito è 0.

Limite di tempo (minuti)
Il numero di minuti consentiti per i calcoli del modello da eseguire. Se si specifica 0, il timer viene disattivato. Il valore predefinito è 5.

Visualizza

Specifica la quantità di output da visualizzare per la modalità **Selezione alfa tramite convalida incrociata** .

Cordiali saluti
Visualizza solo i risultati di base per il modello migliore scelto. È l'impostazione predefinita.

Confronta modelli
Visualizza i risultati di base per tutti i modelli valutati.

Confronta modelli e occorrenze
Visualizza risultati completi verbose per ogni suddivisione o piega per ogni modello valutato.

Grafico

Specifica i grafici dei valori osservati o residui rispetto ai valori previsti e, con la convalida incrociata, la specifica dei grafici dell'errore quadratico medio (MSE) e / o della media R^2 rispetto alle occorrenze di convalida incrociata rispetto ai valori alfa.

MSE (mean squared error) medio di convalida incrociata rispetto a alfa
Per la modalit ... **Selezione alfa tramite convalida incrociata** , visualizza un grafico a linee dell'MSE medio sulle occorrenze di convalida incrociata rispetto all'alfa. Per la modalità **Traccia traccia** , una trama simile viene prodotta automaticamente in base ai dati di formazione completi.

Valore medio di R - quadrato con convalida incrociata rispetto a alfa
Per la modalit ... **Selezione alfa tramite convalida incrociata** , visualizza un grafico a linee della media R^2 sulle occorrenze di convalida incrociata rispetto all'alfa. Per la modalità **Traccia traccia** , una trama simile viene prodotta automaticamente in base ai dati di formazione completi.

Osservato rispetto a Previsto
Visualizza un grafico a dispersione dei valori osservati rispetto a quelli previsti per il modello specificato o migliore.

Residui rispetto a Previsto
Visualizza un grafico a dispersione dei residui rispetto ai valori previsti per il modello specificato o migliore.

Salva

Specifica le variabili da salvare al dataset attivo.

Valori previsti
Salvare i valori previsti dal modello specificato o migliore al dataset attivo. È anche possibile specificare un **Nome variabile personalizzata**.

Residui
Salvare i residui dalle previsioni del modello specificate o migliori al dataset attivo. È anche possibile specificare un **Nome variabile personalizzata**.

Regressione Lineare Ridge

Linear Ridge utilizza la classe Python sklearn.linear_model.Ridge per stimare i modelli di regressione lineare regolarizzata con perdita quadrata o L2 per una variabile dipendente su una o più variabili indipendenti e include modalità facoltative per visualizzare i grafici di traccia e per selezionare il valore dell'iperparametro alfa in base alla convalida incrociata. Quando viene montato un modello singolo o un crossvalidation viene utilizzato per selezionare alpha, una partizione di dati holdout può essere utilizzata per stimare le prestazioni fuori campione.

Oltre ad allestimento di un modello con un valore specificato del parametro di regolarizzazione alfa, la cresta lineare può visualizzare una traccia di traccia dei valori di coefficiente per una gamma di valori alfa, oppure facilitare la scelta del valore iperparametro tramite crossvalidation di k - fold su specifiche griglie di valori. Se viene adattato un singolo modello o viene eseguita la selezione alfa mediante convalida incrociata, il modello finale può essere applicato ai dati trattenuti creati da una partizione dei dati di input per ottenere una stima valida delle prestazioni fuori campione del modello.

Ottenimento di un'analisi Linear Ridge Regression

1. Dai menu, scegliere:

   Analizzare > **Regressione** > **Linear OLS Alternative** > **Ridge**

   La finestra di dialogo consente di specificare una variabile che assegna ogni caso nel dataset attivo al campione di formazione o holdout.
2. Selezionare una variabile di destinazione numerica. È necessaria una sola variabile di destinazione per eseguire un'analisi.
3. Specificare un dipendente numerico.
4. Specificare almeno una variabile di covariata variabile o numerica.

Facoltativamente, **Partizione** fornisce un modo per creare un sottoinsieme di holdout o di test dei dati di input per la stima delle prestazioni fuori campione del modello specificato o scelto. Tutto il partizionamento viene eseguito dopo la cancellazione listica di eventuali casi con dati non validi per qualsiasi variabile utilizzata dalla procedura. Si noti che per il crossvalidation, le pieghe o le partizioni dei dati formativi vengono create in Python. I dati di holdout creati dalla partizione non vengono utilizzati nella stima, indipendentemente dalla modalità in vigore.

La partizione può essere definita specificando il rapporto dei casi assegnati casualmente a ciascun campione (in **Partizioni di addestramento e di holdout**) oppure mediante una variabile che assegna ciascun caso al campione di addestramento o di holdout. Non è possibile specificare sia la formazione che le variabili. Se la partizione non viene specificata, viene creato un campione holdout di circa il 30% dei dati di input.

Il **Formazione%** specifica il numero relativo di casi nel dataset attivo per assegnare casualmente al campione di formazione. La formazione predefinita è di 70%.

Regressione Lineare Ridge: Opzioni

La scheda Opzioni fornisce opzioni per:

Modalità

Questa selezione fornisce opzioni per specificare una delle seguenti modalità:

Adatta con alfa specificato
Quando si seleziona questa opzione, un singolo modello viene adattato ai dati di addestramento che utilizzano solo un valore di regolarizzazione alfa. È impostato per impostazione predefinita. Se viene specificata una partizione, il modello singolo o finale adattato viene applicato ai dati di test tenuti fuori per stimare le prestazioni fuori campione.

In **Grafico**, è possibile selezionare i grafici dei valori osservati e / o residui rispetto ai valori previsti.

Sotto **Salva**è possibile specificare valori e residui previsti da salvare.

Grafico delle tracce
Quando si seleziona questa opzione, vengono visualizzati tre grafici per i dati di addestramento come una funzione di alfa per la serie specificata di valori alfa:

- Una traccia di traccia di coefficienti di regressione.
- Un grafico di R^2.
- Una trama di errore quadrato medio (MSE).

Sebbene la partizione sia onorata, non vengono forniti risultati per i dati di test detenuti perché nessun modello finale risulta da questa modalità.

Selezione alfa tramite convalida incrociata

Quando si seleziona questa opzione, viene effettuata una ricerca a griglia con convalida incrociata per valutare i modelli e viene scelto il migliore alfa in base alla migliore media R^2 rispetto alle occorrenze di convalida. Il campo **Numero di occorrenze di convalida incrociata** può essere utilizzato per modificare il valore predefinito di cinque suddivisioni o occorrenze per la convalida incrociata. Se viene specificata una partizione, il modello singolo o finale adattato viene applicato ai dati di test tenuti fuori per stimare le prestazioni fuori campione.

In **Visualizzazione**, è possibile selezionare la visualizzazione delle informazioni di base solo sul modello con il valore scelto di alfa (**Migliore**), delle informazioni di base su tutti i modelli confrontati (**Confronta modelli**) o delle informazioni complete su tutte le suddivisioni o le occorrenze per tutti i modelli (**Confronta modelli e occorrenze**). **Best** è il default.

In **Grafico**, è possibile selezionare i grafici della media R^2 e / o MSE sulle occorrenze di convalida. È possibile selezionare anche i grafici dei residui osservati e / o rispetto ai valori previsti.

Sotto **Salva**è possibile specificare valori e residui previsti da salvare.

Specifica singoli valori alfa
Quando si seleziona la modalità **Adatta con alfa specificato** , è possibile specificare un singolo valore di regolarizzazione alfa. Quando si seleziona la modalità **Traccia del grafico** o **Selezione alfa tramite convalida incrociata** , è possibile specificare più valori.

Valori
Specificare uno o più valori di regolarizzazione alfa positivi. È possibile specificare più valori singolarmente o come intervalli. Il valore predefinito è 1.

Specifica griglia di valori alfa
Quando si seleziona la modalità **Traccia** o **Selezione alfa tramite convalida incrociata** , è possibile specificare una griglia di valori alfa univoci da un valore **Inizio** (`value1`) a un valore **Fine** (`value2`) con l'incremento di **Da** (`value3`). Se specificato, è consentita una sola serie valida di [`value1 TO value2 BY value3`]. Deve soddisfare tale $0 \leq value1 \leq value2 \leq 1$. Nei casi in cui `value1 = value2`, equivale a specificare un singolo `value1`, indipendentemente da `value3`.

La **metrica Alpha** per intervalli di valori può essere **Linear** o **Base 10 logaritmica** (10 viene innalzato alla potenza dei valori specificati).

I grafici vengono visualizzati utilizzando la metrica specificata per gli assi X orizzontali dei valori alfa variabili.

Criteri

Analisi dei controlli.

Includi intercettazione
Questo criterio include un'intercettazione in uno o più modelli adattati. Si noti che la procedura di estensione non centro o standardizza la variabile dipendente e l'intercettazione non è penalizzata durante la stima.

Standardizza predittori
Questo criterio standardizza tutte le variabili indipendenti.

Numero di occorrenze di convalida incrociata
Utilizzare questo criterio per impostare il numero di suddivisioni o di occorrenze per la valutazione della convalida incrociata dei modelli. Il numero deve essere un valore intero positivo maggiore di 1. L'impostazione predefinita è 5.

Python stato casuale
Il valore dell'impostazione random_state in Python viene utilizzato durante l'esecuzione della valutazione di convalida incrociata dei modelli. Consente la replica dei risultati che coinvolgono numeri pseudo - casuali. Deve essere un intero compreso nell'intervallo da 0 a 2^{32}-1. Il valore predefinito è 0.

Limite di tempo (minuti)
Il numero di minuti consentiti per i calcoli del modello da eseguire. Se si specifica 0, il timer viene disattivato. Il valore predefinito è 5.

Visualizza

Questa sezione specifica la quantità di output da visualizzare per la modalità **Selezione alfa tramite convalida incrociata** .

Cordiali saluti
Visualizza solo i risultati di base per il modello migliore scelto. È impostato per impostazione predefinita.

Confronta modelli
Visualizza anche i risultati di base per tutti i modelli valutati.

Confronta modelli e occorrenze
Infine, visualizza i risultati completi verbose per ogni suddivisione o piega per ogni modello valutato.

Grafico

Specifica i grafici dei valori osservati o residui rispetto ai valori previsti, con convalida incrociata, specifica dei grafici dell'errore quadratico medio (MSE) e / o della media R^2 rispetto alle occorrenze di convalida incrociata rispetto ai valori alfa.

MSE (mean squared error) medio di convalida incrociata rispetto a alfa
La modalità **Selezione alfa tramite convalida incrociata** visualizza un grafico a linee dell'MSE medio rispetto alle occorrenze di convalida incrociata rispetto all'alfa. Per la modalità **Traccia traccia** , una trama simile viene prodotta automaticamente in base ai dati di formazione completi.

Valore medio di R - quadrato con convalida incrociata rispetto a alfa
La modalità **Selezione alfa tramite convalida incrociata** , visualizza un grafico a linee della media R^2 rispetto alle occorrenze di convalida incrociata rispetto all'alfa. Per la modalità **Traccia traccia** , una trama simile viene prodotta automaticamente in base ai dati di formazione completi.

Osservato rispetto a Previsto
Visualizza un grafico a dispersione dei valori osservati rispetto a quelli previsti per il modello specificato o migliore.

Residui rispetto a Previsto
Visualizza anche un grafico a dispersione dei residui rispetto ai valori previsti per il modello specificato o migliore.

Salva

Specifica le variabili da salvare al dataset attivo.

Valori previsti
Salvare i valori previsti dal modello specificato o migliore al dataset attivo. In alternativa, è possibile specificare un **Nome variabile personalizzata**.

Residui

Salvare i residui dalle previsioni del modello specificate o migliori al dataset attivo. In alternativa, è possibile specificare un **Nome variabile personalizzata**.

Analisi di mediazione

L'analisi di mediazione è un metodo statistico utilizzato per comprendere il meccanismo attraverso il quale una variabile indipendente influenza una variabile dipendente tramite una o più variabili mediatrici. Questa procedura ti aiuta a verificare le ipotesi relative agli effetti indiretti e a comprendere i meccanismi di influenza presenti nei tuoi dati.

La procedura "Analisi di mediazione" in IBM SPSS Statistica offre le seguenti funzionalità:

- Analizzare gli effetti diretti e indiretti tra le variabili predittive, mediatrici e di risultato
- Scomporre gli effetti complessivi nelle loro parti costitutive per comprenderne i meccanismi causali
- Gestire vari tipi di variabili, inclusi dati continui, binari e di conteggio
- Controllo delle variabili confondenti tramite aggiustamento per le covariate
- Effettuare analisi per sottogruppi per valutare l'eterogeneità degli effetti
- Visualizza le relazioni di mediazione tramite diagrammi di percorso e altri grafici

Concetti principali

- **Variabile predittiva (A)** : la variabile indipendente di cui si desidera esaminare l'effetto (ad esempio, l'esposizione al trattamento, il livello di intervento).
- **Variabile mediatrice (M)** : la variabile intermedia che trasmette l'effetto della variabile predittiva alla variabile di risultato (ad esempio, stato psicologico, marcatore biologico).
- **Variabile di risultato (Y)** : la variabile dipendente influenzata sia dalla variabile predittiva che dal mediatore (ad esempio, esito sanitario, indicatore di prestazione).
- **Effetto diretto** : l'effetto della variabile predittiva sul risultato che non agisce attraverso il mediatore.
- **Effetto indiretto** : l'effetto della variabile predittiva sul risultato che opera attraverso il mediatore.
- **Effetto totale** : l'effetto complessivo del predittore sul risultato, che tiene conto sia dei percorsi diretti che di quelli indiretti.

Applicazioni

L'analisi di mediazione è ampiamente utilizzata in diverse discipline:

- **Scienze della salute** : comprendere in che modo gli interventi influenzano gli esiti sanitari attraverso meccanismi comportamentali o biologici
- **Psicologia** : Analisi dell'influenza dei trattamenti sulla salute mentale attraverso processi cognitivi o emotivi
- **Scienze sociali** : Analisi dell'impatto delle politiche sui risultati sociali attraverso fattori intermedi
- **Settore aziendale** : Analisi dell'influenza delle strategie di marketing sulle vendite attraverso il comportamento dei clienti
- **Formazione** : studio dell'influenza dei metodi didattici sui risultati dell'apprendimento attraverso il coinvolgimento degli studenti

Questa procedura è particolarmente utile quando è necessario:

- Verificare i modelli teorici relativi ai meccanismi causali
- Identificare gli obiettivi dell'intervento attraverso la comprensione dei meccanismi mediatori
- Valutare l'importanza relativa dei diversi percorsi causali
- Valutare se gli effetti variano tra i sottogruppi
- Controllo delle variabili confondenti negli studi osservazionali

Tipi di analisi

La procedura supporta due approcci di analisi:

- **Analisi di mediazione causale** : calcola gli effetti causali complessivi, tra cui l'effetto diretto controllato (CDE), l'effetto diretto naturale (NDE), l'effetto indiretto naturale (NIE) e le componenti della scomposizione. Questo approccio tiene conto delle interazioni tra variabili predittive e mediatrici e fornisce una scomposizione dettagliata degli effetti.
- **Analisi di mediazione tradizionale** : calcola effetti di mediazione più semplici senza termini di interazione né scomposizioni dettagliate. Questo approccio è indicato quando non si prevedono interazioni tra la variabile predittiva e quella mediatrice.

Combinazioni di variabili supportate

La procedura supporta diverse combinazioni di tipi di variabili:

- Variabile predittiva continua, mediatore continuo, esito continuo
- Variabile predittiva continua, mediatore binario, esito continuo
- Variabile predittiva continua, mediatore continuo, esito binario
- Variabile predittiva continua, mediatore binario, esito binario
- Variabile predittiva continua, mediatore continuo, variabile di risultato discreta

Statistical Methods

La procedura utilizza approcci basati sulla regressione per stimare gli effetti di mediazione:

- Regressione lineare per variabili continue
- Regressione logistica per variabili binarie
- Regressione di Poisson o binomiale negativa per variabili di conteggio

Gli errori standard e gli intervalli di confidenza possono essere calcolati utilizzando il metodo delta (impostazione predefinita) o il ricampionamento bootstrap. Il metodo delta garantisce un calcolo veloce basato sulla teoria asintotica, mentre il bootstrapping offre un'inferenza più robusta per distribuzioni non normali o campioni di piccole dimensioni.

Statistiche

La procedura fornisce risultati statistici completi, suddivisi in diverse categorie:

Sintesi dei modelli di regressione

Modello di regressione della mediazione:

- Coefficienti (β_0, β_1, β_2) con errori standard e valori p
- Per i mediatori continui: $E(M|a,c) = \beta_0 + \beta_1 a + \beta_2 c$
- Per i mediatori binari: modello di regressione logistica

Modello di regressione dei risultati:

- Coefficienti (θ_0, θ_1, θ_2, θ_3, θ_4) con errori standard e valori p
- Per le variabili continue: $E(Y|a,m,c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 (a \times m) + \theta_4 c$
- Per risultati binari: modello di regressione logistica
- Per i risultati con conteggio: regressione di Poisson o regressione binomiale negativa

Misure di effetto

Statistiche comparative sui livelli di esposizione:

- **CDE (effetto diretto controllato)** : effetto della variabile predittiva sul risultato a un livello fisso del mediatore
- **NDE (Effetto diretto naturale)** : effetto diretto che consente al mediatore di variare in modo naturale
- **NIE (Effetto indiretto naturale)** : effetto mediato dal mediatore

- Presentati come differenze per gli esiti continui, come odds ratio (OR) o come excRR per gli esiti binari

Risultati dell'analisi di decomposizione:

- **TE (Effetto totale)** : effetto complessivo della variabile predittiva sul risultato
- **PIE (effetto indiretto puro)** : componente dell'effetto indiretto
- **TDE (Effetto diretto totale)** : componente dell'effetto diretto
- **PAI (Pure Interaction)** : Interazione attribuibile al predittore
- **INTmed (Interazione mediata)** : Interazione tramite un mediatore
- **INTref (Interazione di riferimento)** : effetto di interazione di base

Statistiche di rapporto (per risultati binari):

- **PM (Proporzione mediata)** : Proporzione dell'effetto totale mediata
- **Percentuale (INT)** : Percentuale dovuta all'interazione
- **Percentuale (ELIM)** : Percentuale eliminata

Metodi di inferenza

Metodo Delta (impostazione predefinita):

- Errori standard calcolati dalle matrici di covarianza (Σβ, Σθ)
- Intervalli di confidenza basati sulla distribuzione normale
- Calcolo veloce, adatto alla maggior parte delle analisi

Metodo Bootstrap:

- Errori standard e intervalli di confidenza basati sul ricampionamento
- Impostazione predefinita: 1000 campioni di bootstrap con un livello di confidenza del 95%
- Intervalli percentili o BCa (con correzione del bias e accelerati)
- Maggiore robustezza per distribuzioni non normali

Opzioni della matrice di covarianza:

- **MODELLO** : utilizza la varianza prevista del modello (risultato ottimale quando i dati si adattano bene al modello)
- **ROBUST** : tiene conto dei valori anomali e dell'eteroschedasticità (consigliato per le variabili binarie)

Opzioni di visualzzazione

Diagramma di percorso:

- Rappresentazione grafica dei percorsi di mediazione
- Mostra gli effetti diretti e indiretti con le dimensioni dell'effetto
- Visualizzazione facoltativa degli effetti delle covariate

Diagramma a dispersione:

- Relazioni tra variabile predittiva, mediatrice e risultato
- Linee di tendenza (lineari o logistiche) con intervalli di confidenza

Grafico a barre:

- Conteggi o percentuali per le variabili categoriali
- Barre di errore e intervalli di confidenza

Forest Plot:

- Riepilogo delle misure di effetto con intervalli di confidenza
- Visualizza i coefficienti di percorso o le stime degli effetti

Considerazioni sui dati

Tabella 1. Requisiti variabili

Variabile	Requisito
Variabile risultato	• Dati continui, binari o numerici • Per i risultati binari: specificare la categoria di riferimento (ULTIMO, PRIMO o un VALORE specifico) • Per i risultati numerici: scegliere tra la distribuzione di Poisson e quella binomiale negativa
Variabile mediatrice	• Dati continui o binari • Per i mediatori continui: specificare il livello da analizzare (MEDIA, MEDIANA o un VALORE specifico) • Per i mediatori binari: specificare la categoria di riferimento
Variabile predittiva	• Dati continui o categoriali • Specificare il livello di riferimento (BLEVEL) e il livello di esposizione (ELEVEL) per il confronto degli effetti • Termine opzionale di interazione predittore-mediatore (INTERACTION = TRUE/FALSE)

Specifiche modello

La procedura si avvale del quadro dei modelli lineari generalizzati (GLM), che richiede la specificazione sia della distribuzione che della funzione di collegamento.

Per il modello Mediator	Per il modello dei risultati
• **Una distribuzione normale con funzione di collegamento identitaria** implica una regressione lineare per i mediatori continui • **La distribuzione binomiale con funzione di collegamento logit** implica una regressione logistica per i mediatori binari	• **Una distribuzione normale con funzione di collegamento identitaria** implica una regressione lineare per i risultati continui • **La distribuzione binomiale con funzione di collegamento logit** implica una regressione logistica per gli esiti binari • **La distribuzione di Poisson con funzione di collegamento logaritmica** implica una regressione di Poisson per i dati di conteggio (presupponendo che la varianza sia uguale alla media) • **Distribuzione binomiale negativa con collegamento logaritmico** per dati di conteggio sovradispersi (la varianza supera la media) – **NEGBIN(1)** : Parametro di dispersione fisso $\alpha = 1$ – **NEGBIN(MLE)** : calcola il parametro di dispersione dai dati

Il parametro di dispersione α controlla la sovradispersione: $Var(Y) = \mu + \alpha\mu^2$. Quando $\alpha = 0$, il modello si riduce a quello di Poisson. Valori di α più elevati indicano una maggiore variabilità rispetto alla media.

Dimensione del campione e convergenza

Tabella 2. Parametri degli algoritmi

Parametro	Descrizione	Intervallo
MASSIMO NUMERO DI ITERAZIONI	Numero massimo di iterazioni	da 10 a 1000 Aumentare a 200-500 per i modelli complessi
PCONVERGE	Tolleranza convergenza	1E-8 all'indirizzo 1E-3 Utilizza 1E-8 per le pubblicazioni e 1E-4 per le analisi esplorative
Singolare	Soglia di rilevamento della singolarità	valore numerico compreso tra 1E-14 e 1E-6
MAXSTEPHALVING	Tentativi di dimezzamento graduale	da 1 a 20
PENALITÀ DI RIDGE	Regolarizzazione dei crinali	1E-8 all'indirizzo 1E-3

Considerazioni sulla qualità dei dati

Dati mancanti

- I casi in cui mancano valori nelle variabili di analisi vengono esclusi
- Riepilogo dell'elaborazione dei casi: casi validi ed esclusi

Pesi

- I pesi di frequenza sono supportati tramite la variabile "Scale Weight"
- Il ricampionamento Bootstrap arrotonda i pesi al numero intero più vicino

Covariate

- Includere le variabili covarianti per controllare i fattori di confondimento
- Tenere conto dei fattori che potrebbero influenzare sia il mediatore che l'esito

Analisi del sottogruppo

- Utilizzare la clausola BY per eseguire analisi separate in base a determinati fattori (ad esempio, sesso, fascia d'età)
- Utile per esaminare l'eterogeneità degli effetti tra i gruppi

Ipotesi

Ipotesi del modello

- Non sono stati riscontrati fattori di confondimento non misurati nelle relazioni tra predittore e mediatore, tra predittore e esito e tra mediatore ed esito
- Corretta definizione delle funzioni di distribuzione e di collegamento
- Per l'interpretazione causale: sequenza temporale (predittore → mediatore → esito)
- Nessun errore di misurazione nelle variabili

Ipotesi statistiche

- Dimensione adeguata del campione per ottenere stime dei parametri stabili
- Per il bootstrap: un numero sufficiente di campioni (si consiglia un minimo di 1000)
- Per gli esiti binari: numero adeguato di eventi per variabile (si consiglia un minimo di 10-15)
- Per i modelli di interazione: potenza statistica sufficiente per rilevare gli effetti di interazione

Ottenere un'analisi di mediazione

1. Dai menu, scegliere:

   Analisi > **Analisi di mediazione**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

Tipo di modello

Nella scheda "**Tipo di modello** ", definire la natura statistica dell'analisi di mediazione selezionando il " **Tipo di mediatore**" e **il "Tipo di risultato"**.

Seleziona il tipo di modello di mediazione che desideri analizzare. Le impostazioni nei campi **"Tipo di mediatore"** e **"Tipo di risultato"** determinano la distribuzione sottostante, la funzione di collegamento e quali variabili sono consentite nelle schede successive. Le impostazioni selezionate in questa scheda determinano quali opzioni sono attive nelle schede **"Variabili dipendenti"** e **"Criteri"**.

I campi "Tipo di mediatore" e "Tipo di risultato" sono obbligatori.

Tipo di mediatore

Selezionare il modo in cui viene modellata la variabile mediatrice.

Continuo
Si ipotizza una distribuzione normale con una funzione di collegamento identitaria. Per il mediatore sono ammesse solo variabili continue.

Utilizza questa opzione quando il mediatore viene misurato su una scala continua.

Binario

Si ipotizza una distribuzione binomiale con funzione di collegamento LOGIT. Per il mediatore sono ammesse solo variabili binarie.

Utilizza questa opzione quando il mediatore può assumere due valori diversi.

Tipo di risultato

Scegliere come modellare la variabile di risultato. Le tabelle di output vengono generate quando la prevalenza della variabile di risultato è pari ad almeno l' 0.30 e.

Continuo
Si ipotizza una distribuzione normale con una funzione di collegamento identitaria.

Selezionare l'opzione per impostare il tipo di variabile a valori continui, il che implica una distribuzione NORMALE dei dati e una funzione di collegamento IDENTITY. Se selezioni questa opzione, puoi configurare la **variabile di risultato** nella scheda **"Variabili dipendenti"** in modo da consentire solo variabili continue.

Binario

Si ipotizza una distribuzione binomiale. È possibile selezionare la funzione **LOGIT** o **LOG**.

Selezionare l'opzione per impostare il tipo di risultato binario con distribuzione BINOMIALE e funzione di collegamento LOGIT. Se selezioni questa opzione, puoi configurare il campo **"Variabile di risultato"** nella scheda **"Variabili dipendenti"** e selezionare "Binario" come tipo di risultato.

Dall'elenco a discesa "Funzione di collegamento" è possibile selezionare "**Logit** " (impostazione predefinita) o "Funzione di collegamento **logaritmica** per esito binario". Per impostazione predefinita, l'opzione **Binary** imposta la distribuzione BINOMIALE e la funzione di collegamento LOGIT. Nella scheda "**Variabili dipendenti** ", è possibile specificare la variabile di risposta come **"Binaria** con **categoria di riferimento"** per selezionare tra valori binari.

Conteggi
Selezionare questa opzione per configurare la variabile «counts» come **«Outcome»** nella scheda «**Variabili dipendenti** ». È possibile aggiungere distribuzioni **di Poisson** o **binomiali negative** con una funzione di collegamento logaritmica. Per l'opzione **"Binomiale negativo"** sono disponibili ulteriori impostazioni dei parametri. È possibile impostare il parametro di dispersione α che controlla il grado di sovradispersione nel modello. Valori più elevati di α indicano una maggiore variabilità rispetto alla media.

- **Specificare il valore** : inserire un numero intero positivo fisso.
- **Valore stimato** : selezionare questa opzione se si desidera stimare il parametro di dispersione α utilizzando il metodo della massima verosimiglianza.

Variabili dipendenti

Nella scheda " **Variabili dipendenti** " è possibile specificare la variabile di risultato, che definisce l'esito dell'analisi di mediazione.

Specificare la variabile di risposta (risultato), con opzioni di personalizzazione per variabili continue, binarie (compresa la categoria di riferimento) e di conteggio. Inoltre, definisce la variabile mediatrice con opzioni per variabili continue e binarie (compresa la categoria di riferimento). È possibile impostare i livelli di controllo per l'analisi degli effetti e per l'inclusione di una variabile di ponderazione della scala.

È inoltre possibile specificare la variabile Mediator nel campo " **Variabile Mediator** ", che definisce il componente mediatore per l'analisi di mediazione. Tutte le variabili devono essere univoche sia all'interno delle variabili dipendenti e indipendenti, sia rispetto alla variabile "Peso della scala".

Variabile risultato
Definisce l'outcome per l'analisi di mediazione. Fornire variabili continue, binarie e di conteggio. Ciò dipende dal tipo di risultato specificato nella scheda "**Tipo di modello** "; l'impostazione predefinita è "variabile continua". Se si opta per un risultato binario, è possibile selezionare variabili binarie. Se si sceglie la distribuzione di Poisson o la distribuzione binomiale negativa, è possibile selezionare anche variabili di conteggio.

Variabile con esito binario
È possibile selezionare "binario" come classe di risposta per la regressione con classe di risposta. Se si seleziona un risultato binario, è possibile selezionare il valore di riferimento per la variabile binaria selezionata. Per impostazione predefinita, una distribuzione BINOMIALE e una funzione di collegamento LOGIT/LOG.

Le tabelle di output vengono generate quando la prevalenza della variabile di risultato è pari ad almeno l' 0.30 e.

Categoria di riferimento
Con questa opzione è possibile definire la categoria di riferimento per il tipo di risultato binario selezionato nella scheda "**Tipo di modello** ". Scegli una delle seguenti opzioni:

- **Ultimo (valore massimo)** : imposta l'ultimo valore o il valore massimo come riferimento.
- **Primo (valore più basso)** : impostare il primo valore o quello più basso come riferimento.
- **Personalizzato** : selezionare questa opzione per impostare una selezione di riferimento personalizzata. È possibile selezionare i livelli della variabile binaria selezionata nella variabile di risultato oppure inserire un valore, a condizione che tale valore sia presente nell'elenco delle variabili.

Variabile mediatrice
Con questa opzione è possibile definire la **variabile mediatrice** per l'analisi di mediazione. Il tipo di dati deve essere continuo e binario.

Questa opzione si basa sul **tipo** di mediatore specificato nella scheda "**Tipo di modello** "; l'impostazione predefinita prevede una variabile continua, una distribuzione normale e la funzione di collegamento IDENTITY. Se si sceglie un mediatore binario, è possibile selezionare variabili binarie con una distribuzione binomiale e la funzione di collegamento LOGIT.

Imposta il livello del mediatore
È possibile selezionare un livello di mediatore da applicare al calcolo dell'effetto diretto controllato. Scegli una delle seguenti opzioni:

- **Media** : selezionare questa opzione per impostare il livello del mediatore sulla media della variabile mediatrice selezionata.
- **Mediana** : selezionare questa opzione per impostare il livello del mediatore sulla mediana della variabile mediatrice selezionata.
- **Personalizzato** : selezionare questa opzione per specificare un livello personalizzato. Questa opzione può essere selezionata solo per le variabili binarie. Inserisci un valore numerico per il livello del mediatore. Il valore deve essere presente nei valori osservati del file binario Mediator. Se il mediatore è binario, l'utente può selezionare il livello binario (0 o 1).

Peso scala
Seleziona l'opzione per assegnare una variabile di ponderazione della scala. Il campo ammette solo variabili con tipo di dati numerico.

Predittore

Nella scheda "**Predittore** " è possibile specificare la variabile predittiva nel campo "Variabile **predittiva primaria** ", che definisce la componente **predittiva** o **di esposizione**, indicata con *la lettera A*, nell'analisi di mediazione.

Specificare la variabile predittiva principale insieme a fattori e covariate aggiuntivi, con la possibilità di impostare i valori di riferimento e di esposizione per la variabile predittiva al fine di facilitare l'analisi degli effetti.

I valori osservati della variabile predittiva nel set di dati sono rappresentati dal vettore a.

Predittore primario
Definisce la variabile predittiva o di esposizione primaria A per l'analisi di mediazione.

Sono ammesse solo variabili continue (le variabili categoriali sono escluse per garantire chiarezza nella definizione dell'effetto dell'esposizione).

Imposta il livello della linea di base
È possibile specificare i livelli di riferimento utilizzando questa opzione. Inserire i valori nei campi: "**Imposta livello di riferimento** " e "**Imposta nuovo livello di esposizione** ". Entrambi i valori devono essere presenti nel set di dati per la variabile selezionata come predittore.

- Imposta livello di riferimento: il livello più basso (o il primo valore) della variabile predittiva nel suo ordine naturale.
- Imposta nuovo livello di esposizione: il livello immediatamente superiore (o il secondo valore) della variabile predittiva nello stesso ordine.

Fattori
Questo elenco definisce i fattori F che rappresentano le variabili categoriali nel modello. Ogni fattore viene convertito in una serie di variabili dummy da includere nella matrice C. Queste variabili dummy vengono incluse nella regressione dei risultati per tenere conto delle differenze tra i gruppi.

Sono ammesse variabili categoriali (ad esempio, gruppi di trattamento, sesso e così via).

I fattori sono facoltativi, ma se l'analisi prevede gruppi categoriali o se si valutano effetti di interazione, queste variabili sono obbligatorie.

Covariate

Questo elenco definisce le covariate *C* che tengono conto dei potenziali effetti di confondimento nel modello. Ogni covariata è inclusa nel modello di regressione dell'esito Y.

Se le variabili covariate sono categoriali, devono essere convertite in variabili dummy prima di essere incluse nel modello, al fine di evitare ridondanze.

Ogni covariata viene inclusa come effetto principale nel modello di regressione dell'esito Y.

Criteri

Nella scheda **Criteri** è possibile impostare le opzioni relative alla scelta dei metodi di stima per la matrice delle covariate, alla costruzione delle stime dei parametri e alla specificazione dei tipi di inferenza, del tipo di analisi e delle opzioni per la regolazione dell'algoritmo.

Matrice di covarianza

Determina il metodo utilizzato per stimare la matrice di covarianza delle stime dei parametri nell'analisi di mediazione.

- **Stimatore basato su modello** : calcola la matrice di covarianza in base alle ipotesi del modello statistico specificato.
- **Stimatore robusto** : fornisce una stima robusta della matrice di covarianza che tiene conto delle violazioni del modello, quali l'eteroschedasticità o la non normalità.

Tipo di inferenza

Questa opzione serve a specificare il metodo utilizzato per l'inferenza nell'analisi di mediazione.

- **Delta** : Questo metodo utilizza la distribuzione asintotica dello stimatore per calcolare gli errori standard e gli intervalli di confidenza.
- **Bootstrap** : questo metodo utilizza il ricampionamento per ottenere una distribuzione dello stimatore ai fini dell'inferenza, compresi gli intervalli di confidenza.

  Opzioni:

  – **Numero di campioni** : per gli intervalli percentili e BCa generati, si raccomanda di utilizzare almeno 1000 campioni di bootstrap. Specificare un intero positivo. Il valore predefinito è 500.

  – **Imposta il seme per Mersenne Twister** : l'impostazione di un seme consente di ripetere le analisi. L'utilizzo di questo controllo è simile all'impostazione di Mersenne Twister come generatore attivo e alla specifica di un punto iniziale fisso nella finestra di dialogo Generatori di numeri casuali, con la differenza importante che impostando il seme in questa finestra di dialogo preserverà lo stato corrente del generatore di numeri casuali e ripristinerà tale stato una volta completata l'analisi.

  – **Intervalli di confidenza** : specificare un livello di confidenza compreso tra 50 e 100. Gli intervalli percentili utilizzano semplicemente i valori bootstrap ordinati corrispondenti ai percentili dell'intervallo di confidenza. Ad esempio, un intervallo di confidenza al 95° percentile utilizza i percentili 2.5th e 97.5th dei valori di bootstrap come limiti inferiore e superiore dell'intervallo (interpolando i valori di bootstrap se necessario). Gli intervalli con correzione del bias e accelerati (BCa) sono intervalli corretti che risultano più precisi, ma richiedono più tempo per il calcolo.

Tipo di analisi

- **Causale** : calcola gli effetti di mediazione causale completi, inclusi CDE, NDE, NIE, effetti totali, odds ratio, rischi relativi in eccesso e decomposizioni di interazione.
- **Tradizionale** : calcola effetti di mediazione più semplici senza termini di interazione né una decomposizione causale completa.

Opzioni algoritmo

Le opzioni sono attive quando, nella scheda "**Tipo di modello** ", il "Tipo di risultato" è impostato su "**Binario** " o **"Conteggio"**.

- **Numero massimo di iterazioni** : imposta il numero massimo di iterazioni IRWLS prima che l'algoritmo si interrompa. È possibile inserire valori interi compresi tra 10 e and1000. Il valore predefinito è 100.
- **Tolleranza di convergenza** : imposta la soglia per la convergenza dei parametri (variazione assoluta della norma del vettore dei coefficienti).

  Soglia di convergenza dei parametri. Il valore predefinito è 1E-6. L'algoritmo si arresta quando la variazione assoluta dei coefficienti scende al di sotto di questo valore. Utilizza 1E-8 per ottenere una precisione degna di una pubblicazione e 1E-4 per analisi esplorative più rapide.

  Inserisci un valore numerico compreso tra 1E-8 e 1E-3.

  Il campo accetta i seguenti formati in notazione scientifica: 1E-6, 1e-6, 0.000001
- **Tolleranza di singolarità** : imposta la soglia per il rilevamento delle matrici Hessiane singolari (soglia del determinante). Soglia per il rilevamento di matrici Hessiane singolari. Il valore predefinito è 1E-12. L'algoritmo rileva una quasi-singolarità quando il determinante dell'Hessiano scende al di sotto di questo valore. Utilizza 1E-10 per confrontare i risultati di R e Stata. Si vedano i seguenti prerequisiti per la tolleranza alla singolarità:
  - Valori più bassi indicano un approccio più prudente (meno rilevamenti errati di singolarità).
  - Inserisci un valore numerico compreso tra 1E-14 e 1E-6
  - Il campo accetta i seguenti formati in notazione scientifica: 1E-12, 1e-12, 0.000000000001
- **Riduzione della lunghezza massima dei passi** :

  Questo campo indica il numero massimo di tentativi di dimezzamento del passo per ogni iterazione. Il valore predefinito è 5. Il metodo dello step-halving garantisce che la log-verosimiglianza aumenti ad ogni iterazione. Se si verificano errori di dimezzamento dei passi, aumentare il valore a 10-20. Il valore deve essere un numero intero compreso tra 1 e 20.
- **Penalità di Ridge** : imposta l'intensità della regolarizzazione di Ridge per le matrici Hessiane singolari.

  Il valore predefinito è 1E-6. Quando la matrice Hessiana è quasi singolare (determinante inferiore alla tolleranza di singolarità), viene applicata questa penalità per stabilizzare l'inversione della matrice. Utilizzare valori più elevati (1E-4, 1E-3) per le variabili predittive altamente collineari.

  Inserisci un valore numerico compreso tra 1E-8 e 1E-3. Il campo accetta anche i formati in notazione scientifica: 1E-6, 1e-6, 0.000001

Stampa

Nella scheda **Stampa** è possibile impostare le opzioni per personalizzare l'output in modo da includere solo gli elementi rilevanti selezionati.

Selezionare una delle seguenti opzioni per generare la sezione di output corrispondente come parte dei risultati dell'analisi.

Riepilogo del modello di regressione della mediazione
: Questa opzione è selezionata di default in quanto opzione integrata e non può essere modificata. Nell'output viene generata una tabella che riassume i risultati del modello di regressione mediata. Include i coefficienti stimati (β0, β1, β2), i relativi errori standard e i livelli di significatività statistica. Questi risultati aiutano a comprendere la relazione tra la variabile predittiva (*A*), il mediatore (*M*) e le covariate (*C*). Questa opzione è selezionata per impostazione predefinita.

Riepilogo del modello di regressione del risultato
: Questa opzione è selezionata di default in quanto opzione integrata e non può essere modificata. Nell'output viene generata una tabella che riassume i risultati del modello di regressione. Fornisce i coefficienti stimati (θ0, θ1, θ2, θ3, θ4), i relativi errori standard e i livelli di significatività. Questi risultati valutano gli effetti del predittore, del mediatore e delle covariate sull'esito (*Y*). Questa opzione è selezionata per impostazione predefinita.

Statistiche sul confronto dei livelli di esposizione
Questa opzione è selezionata di default in quanto opzione integrata e non può essere modificata. Nel risultato viene generata una sezione che riporta le differenze nei risultati previsti in base ai livelli di esposizione specificati della variabile predittiva. Le statistiche includono stime quali l'effetto diretto controllato (*CDE*), l'effetto diretto naturale (*NDE*) e l'effetto indiretto naturale (*NIE*), fornendo indicazioni su come le variazioni dell'esposizione influenzano il mediatore e l'esito.

Risultati dell'analisi di decomposizione
Selezionare questa opzione per generare una tabella che presenti la scomposizione dell'effetto totale (*TE*) nelle componenti dirette e indirette, inclusi *CDE*, *NDE* e *NIE*. Fornisce inoltre ulteriori informazioni quali la proporzione mediata (*PM*) e gli effetti di interazione.

Analisi degli effetti di interazione
Selezionare questa opzione per generare una sezione che valuti gli effetti di interazione, quali *INTmed* e *INTref*, che misurano il contributo delle interazioni mediatore-esito e predittore-mediatore. Questi risultati aiutano a comprendere in che modo le interazioni influenzano l'esito.

Statistiche dei rapporti
Selezionare questa opzione per generare statistiche sui rapporti quali *r(CDE)*, *r(NDE)*, *r(NIE)* e *r(INTmed)*. Questi valori rappresentano la percentuale dell'effetto totale (*TE*) attribuibile a ciascuna componente, espressa in percentuale.

Matrice di covarianza per le stime dei parametri
Questa opzione è selezionata per impostazione predefinita. È possibile generare una matrice che riporti le covarianze tra le stime dei parametri, come β (dalla regressione di mediazione) e ϑ (dalla regressione dell'esito). È fondamentale per il calcolo degli errori standard delle statistiche derivate utilizzando il metodo delta. Questa opzione è selezionata per impostazione predefinita.

Riepilogo elaborazione casi
Questa opzione è selezionata per impostazione predefinita. È possibile generare un riepilogo che fornisca una panoramica dei dati utilizzati nell'analisi, compreso il numero di casi, i casi esclusi ed eventuali pesi applicati. Aiuta a valutare la completezza e la qualità del set di dati. Questa opzione è selezionata per impostazione predefinita.

Statistiche descrittive
Selezionare questa opzione per generare una sezione che riporti le statistiche descrittive relative a tutte le variabili coinvolte nell'analisi, quali medie, deviazioni standard, valori minimi e valori massimi. Fornisce una panoramica della distribuzione e della variabilità dei dati.

Grafici

Nella scheda "**Grafici** " è possibile impostare le opzioni per generare visualizzazioni che aiutano a spiegare le relazioni tra la variabile mediatrice, la variabile predittiva e la variabile di risultato.

Sono disponibili diverse opzioni per definire quattro tipi di visualizzazione: diagrammi di percorso, grafici a dispersione, grafici a barre e grafici a foresta. Ogni tipo di grafico offre una prospettiva diversa sul modello di mediazione e può essere personalizzato in base alle proprie esigenze di analisi.

È possibile personalizzare ogni grafico aggiungendo le dimensioni dell'effetto, gli intervalli di confidenza o le suddivisioni per sottogruppi.

Diagramma del percorso
Visualizza le relazioni tra la variabile mediatrice, la variabile predittiva e la variabile di risultato. I diagrammi di percorso sono utili per illustrare il modello di mediazione nel suo complesso, compresi gli effetti diretti e indiretti.

- **Nascondi effetti delle covariate** : nasconde tutti i bordi associati alle covariate. I percorsi predittore-mediatore, mediatore-esito ed esito-predittore rimangono visibili.
- **Colore dei bordi** : quando si seleziona questa opzione, ai bordi viene assegnato un codice colore che indica il tipo di effetto (ad esempio, effetti diretti, indiretti e delle covariate). Il verde indica effetti più intensi, mentre il grigio indica effetti più deboli. Se questa opzione non è selezionata, il colore predefinito del bordo è uniforme.

Grafico a dispersione

I grafici a dispersione mostrano le relazioni tra le variabili e aiutano a individuare le correlazioni e i modelli di distribuzione.

Nell'analisi di mediazione, i grafici a dispersione aiutano a visualizzare le relazioni tra la variabile predittiva (X), il mediatore (M) e la variabile di risultato (Y). Queste relazioni corrispondono ai seguenti percorsi di mediazione:

- Percorso a: Effetto della variabile predittiva (X) sulla variabile mediatrice (M)
- Percorso b: Effetto del mediatore (M) sull'esito (Y)
- Percorso c′: Effetto diretto della variabile predittiva (X) sull'esito (Y)

Un grafico a dispersione a tre variabili mostra come la variabile predittiva influenzi sia il mediatore che l'esito, illustrando il ruolo del mediatore nella relazione tra la variabile predittiva e l'esito. La visualizzazione comprende in genere:

- Punti dati che rappresentano la relazione tra la variabile predittiva e la variabile mediatrice (percorso a)
- Dati che illustrano la relazione tra il mediatore e l'esito (percorso b)
- La relazione complessiva tra la variabile predittiva e l'esito, compresi sia gli effetti diretti che quelli indiretti (percorso c′)

Grazie ai grafici a dispersione, è possibile valutare visivamente l'intensità, la direzione e la significatività di queste relazioni. Gli assi rappresentano solitamente le variabili mediatrici e quelle di risultato, mentre la variabile predittiva viene rappresentata tramite sfumature di colore o indicatori.

Grafico a barre

Mostrano le proporzioni o i conteggi delle variabili categoriali e sono utili per visualizzare i livelli dei mediatori e degli esiti. Questa opzione è disponibile solo quando la variante di mediazione selezionata richiede un grafico a barre (ad esempio, mediatore binario, risultato binario o mediatore continuo con risultato binario).

L'opzione è abilitata per le varianti con mediatore binario, esito binario o mediatore continuo con esito binario.

L'opzione è disabilitata per le varianti con mediatore continuo e risultato continuo, dove vengono utilizzati invece i grafici a dispersione.

- **Barre di errore** : aggiunge le barre di errore ai grafici a barre. Nei grafici a barre raggruppate, le barre di errore rappresentano l'incertezza nelle proporzioni (Y = 1 o Y = 0) per ciascun livello del mediatore.
- **Intervalli di confidenza** : visualizza gli intervalli di confidenza per le barre di errore. Questo valore è determinato dal dato inserito nel campo " **Livello dell'intervallo** di confidenza %" nella scheda **"Criteri"**. Il 95% è il valore predefinito per il livello dell'intervallo di confidenza.

Grafico Forest

Visualizza i coefficienti di percorso e i relativi intervalli di confidenza, che ti aiutano a valutare l'entità e l'incertezza degli effetti di mediazione.

I grafici a foresta mostrano le dimensioni dell'effetto per i percorsi di mediazione (ad esempio, a, b e c′) insieme ai relativi intervalli di confidenza.

Opzioni:

- **Coefficienti di percorso** : visualizza i coefficienti di percorso (a, b e c′) per ciascun percorso di mediazione.
- **Intervalli di confidenza** : visualizza gli intervalli di confidenza per ciascun coefficiente di percorso. Questo valore è determinato dal dato inserito nel campo " **Livello dell'intervallo** di confidenza %" nella scheda **"Criteri"**. Il 95% è il valore predefinito per il livello dell'intervallo di confidenza.

Filtri per serie temporali

Nell'analisi delle serie temporali, è essenziale rimuovere i valori che escono da un intervallo deciso e che causano fluttuazioni insolite nella serie temporale. Un buon filtro deve essere in grado di rimuovere le radici unitarie e le componenti cicliche, oppure deve essere in grado di isolare le fluttuazioni dei dati a una certa frequenza.

Il filtraggio delle serie temporali è una tecnica fondamentale in econometria e nell'analisi delle serie temporali, spesso utilizzata per scomporre una serie temporale in componenti di tendenza e cicliche. Ciò è particolarmente importante in macroeconomia e finanza, dove è essenziale identificare le fluttuazioni del ciclo economico o le tendenze economiche a lungo termine.

I filtri delle serie temporali ampiamente utilizzati sono il filtro Hodrick-Prescott (HP), il filtro Baxter-King (BK) e il filtro Christiano-Fitzgerald (CF).

Ottenere filtri per le serie temporali

I filtri per le serie temporali sono disponibili come procedura di estensione con il plug-in Python. Pertanto, installare il modulo di estensione per eseguire la procedura.

completare i seguenti passaggi per eseguire la procedura per i filtri delle serie temporali.

1. Passare a **DataDefine** > **date and time** per generare le variabili di data.
2. Dopo aver definito la data e l'ora, fare clic su **Analizzare** > **Previsioni** > **Filtri per serie temporali (TSF)**.
3. Selezionare il filtro delle serie temporali che si desidera applicare al set di dati.

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
4. Selezionare le variabili e spostarle nell'elenco delle variabili per il filtro della serie temporale appropriato.
5. Impostare il valore di **Lambda** per il filtro Hodrick-Prescott (HP) o impostare i valori di **Low**, **High** e **K** per i filtri Baxter-King (BK) e Christiano-Fitzgerald (CF).

   Opzionalmente, è possibile selezionare **Deriva** per il filtro Christiano-Fitzgerald (CF).
6. Fare clic su **Esegui**.

Filtri per serie temporali: Opzioni

La finestra di dialogo Filtri serie temporali contiene le seguenti opzioni.

Importante: Prima di impostare le opzioni, andare in **DatiDefinire** > **data e ora** per generare le variabili di data.

Hodrick-Prescott (HP)

Il filtro **Hodrick-Prescott (HP)** è uno strumento utilizzato in macroeconomia per adattare un trend omogeneo alle serie temporali.

Il filtro Hodrick-Prescott separa una serie temporale $_{yt}$ in una componente di tendenza $_{\tau t}$ e in una componente ciclica $_{ct}$.

$y_t = \tau_t + c_t + \square_t$dove $_{\square}$ è la componente di errore (o rumore).

I componenti sono determinati minimizzando la funzione di perdita quadratica e il filtro HP è esplicitamente dato dalla formula,

$HP = [\ ^{\lambda L2-}\ 4\lambda L + (1 + 6\lambda) -\ ^{4\lambda L-1} + \ ^{\lambda L-2}\ ]^{-1}$

dove λ indica il parametro di lisciatura (valore positivo) e L indica l'operatore di ritardo, derivato dalla condizione del primo ordine per il problema di minimizzazione.

Scegliere una variabile per il filtro HP dall'elenco delle variabili e impostare un valore per **Lambda**.

Il parametro di lisciatura λ determina l'equilibrio tra adattamento e lisciatura. Più grande è il valore di λ, più alta è la penalità. Quando si analizzano dati trimestrali, per λ si consiglia il valore predefinito di 1600. Inoltre, λ dovrebbe essere $^{1600/44}$ = 6.25 per i dati annuali e 1600 $*$ ^{34} = 129.600 per i dati mensili. Quando λ= 0, significa che l'uscita della serie di filtri HP corrisponde esattamente alla serie di ingresso. Al contrario, valori più elevati di λ producono serie di output più uniformi.

Baxter-King (BK)

Baxter-King (BK) è un filtro passa-banda comunemente utilizzato nell'analisi delle serie temporali per isolare le componenti cicliche dei dati, spesso negli studi macroeconomici. Il suo obiettivo è quello di rimuovere il rumore ad alta frequenza (a breve termine) e le componenti a bassa frequenza (trend a lungo termine), concentrandosi sui cicli di interesse a medio termine, tipicamente tra 1.5 e 8 anni. Il filtro BK presuppone la stazionarietà delle serie temporali sottostanti. Dall'elenco delle variabili, scegliere 1 o 2 variabili per il filtro BK.

Impostare i valori appropriati in **Basso** (periodo minimo), **Alto** (periodo massimo) e **K** (lunghezza del lead-lag).

Christiano-Fitzgerald (CF)

Il filtro **Christiano-Fitzgerald (CF)** è un altro filtro passa-banda utilizzato per separare una serie temporale in componenti di tendenza e cicliche. Questo filtro è simile al filtro Baxter King (BK), che funziona anch'esso sulla base di un filtro passa-banda ideale, dove il filtro BK assume una media mobile simmetrica per approssimare il filtro passa-banda, ma il filtro CF assume che la serie temporale segua la casualità senza la deriva. Il filtraggio CF viene implementato stimando un filtro ottimale sotto l'ipotesi che i dati siano generati da un processo random walk o di tipo ARIMA.

Dall'elenco delle variabili, scegliere 1 o 2 variabili per il filtro CF. Impostare i valori appropriati in **Low**, **High** e **K**. Selezionare **Drift** se necessario per il filtro CF.

Lambda, **Low** e **High** possono essere valori float qualsiasi. Il valore di **K** deve essere un numero intero. Vedere i seguenti valori:

Basso	1.5 (annuale) e 6 (trimestrale)
Alto	8 (annuale) e 32 (trimestrale)
T	3 (annuali) e 12 (trimestrali)

Stima di curve

La procedura Stima di curve produce le statistiche di regressione per la Stima di curve e i grafici correlati per 11 diversi modelli di regressione per la stima di curve. Per ciascuna variabile dipendente verrà creato un modello distinto. È inoltre possibile salvare come nuove variabili i valori attesi, i residui e gli intervalli di previsione.

Esempio. Un provider di servizi Internet deve tener traccia della percentuale di traffico e-mail infettato da virus sulla propria rete nell'arco di un periodo di tempo specifico. Il grafico di dispersione indica che la relazione non è lineare. È necessario adattare ai dati un modello quadratico o cubico e controllare la validità delle ipotesi e la bontà di adattamento del modello.

Statistiche. Per ogni modello: coefficienti di regressione, *R* multiplo, R^2, R^2 adattato, errore standard della stima, tabella di analisi della varianza, valori previsti, residui e intervalli di previsione. Modelli: lineare, logaritmico, inverso, quadratico, cubico, potenza, composto, curva S, logistico, crescita ed esponenziale.

Considerazioni sui dati relativi alla stima di curve

Dati. Le variabili dipendenti ed indipendenti devono essere quantitative. Se si seleziona **Tempo** come variabile indipenente dal dataset attivo anziché selezionare una variabile, la procedura Stima di curve genera una variabile tempo se il periodo di tempo tra i casi è uniforme. Se si seleziona **Tempo**, la variabile dipendente deve essere una serie storica. L'analisi di serie storiche richiede nel file di dati una struttura in cui ciascun caso (riga) rappresenta un insieme di osservazioni eseguite a orari diversi e il periodo di tempo fra i casi è uniforme.

Ipotesi. Schermate graficamente i vostri dati per determinare come le variabili indipendenti e dipendenti sono correlate (linearmente, esponenzialmente, ecc.). I residui di un buon modello devono essere normali e distribuiti casualmente. Se si utilizza un modello lineare, devono essere soddisfatte le seguenti ipotesi: per ogni valore della variabile indipendente, la distribuzione della variabile dipendente deve essere normale. La varianza della distribuzione della variabile dipendente deve essere costante per tutti i valori della variabile indipendente. La relazione tra la variabile dipendente e la variabile indipendente deve essere lineare e tutte le osservazioni devono essere indipendenti.

Per ottenere una stima di curve

1. Dai menu, scegliere:

   Analizza > **Regressione** > **Stima di curve...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili dipendenti. Per ciascuna variabile dipendente verrà creato un modello distinto.
3. Selezionare una variabile indipendente (una variabile nel dataset attivo oppure **Tempo**).
4. Oppure:

- Selezionare una variabile per etichettare casi nei grafici a dispersione. Per ciascun punto del grafico a dispersione, è possibile utilizzare lo strumento di selezione dei punti per visualizzare il valore della variabile Etichetta di caso.
- Fare clic su **Salva** per salvare i valori attesi, i residui e gli intervalli di previsione come nuove variabili.

Sono inoltre disponibili le seguenti opzioni:

- **Includi costante nell'equazione.** Consente di valutare un termine costante nell'equazione di regressione. La costante viene inclusa per impostazione predefinita.
- **Modelli di grafici.** Consente di tracciare i valori della variabile dipendente e ciascun modello selezionato in base alla variabile indipendente. Viene prodotto un grafico per ogni variabile dipendente.
- **Visualizza tabella ANOVA.** Consente di visualizzare una tabella di analisi della varianza per ciascun modello selezionato.

Stima di curve: Modelli

È possibile scegliere uno o più modelli di regressione per la stima di curve. Per determinare il modello da utilizzare, tracciare i dati in un grafico. Se le variabili appaiono legate da una relazione lineare, utilizzare un modello di regressione lineare semplice. Se la relazione tra le variabili non è lineare, provare a trasformare i dati. Se la trasformazione non risulta utile, può essere necessario utilizzare un modello più complesso. Visualizzare i dati in un grafico a dispersione; se il grafico è simile a una funzione matematica nota, adattare i dati a quel tipo di modello. Se, ad esempio, i dati sono simili a una funzione esponenziale, utilizzare il modello esponenziale.

Lineare. Modello la cui equazione è Y = b0 + (b1 * t). I valori della serie vengono modellati come una funzione lineare del tempo.

Logaritmico.Modello la cui equazione è Y = b0 + (b1 * ln(t)).

Inverso.Modello la cui equazione è Y = b0 + (b1 / t).

Quadratico.Modello la cui equazione è Y = b0 + (b1 * t) + (b2 * t**2). Il modello quadratico può essere usato per modellare una serie che "decolla" o una serie che si smorza rapidamente.

Cubico.Modello definito dall'equazione Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3).

Alimentazione .Model la cui equazione è Y = b0 * (t**b1) o ln(Y) = ln(b0) + (b1 * ln(t)).

Composto.Modello la cui equazione è Y = b0 * (b1**t) o ln(Y) = ln(b0) + (ln(b1) * t).

Curva S.Modello la cui equazione è Y = e**(b0 + (b1/t)) or ln(Y) = b0 + (b1/t).

Logistico.Modello la cui equazione è Y = 1 / (1/u + (b0 * (b1**t))) o ln (1/y-1/u) = ln (b0) + (ln (b1) * t) dove u è il valore del limite superiore. Per usare l'equazione di regressione, specificare il valore limite superiore dopo aver selezionato Logistico. Il valore deve essere un positivo maggiore del valore più alto della variabile dipendente.

Crescita.Modello la cui equazione è Y = e**(b0 + (b1 * t)) o ln(Y) = b0 + (b1 * t).

Exponenziale. Modello la cui equazione è Y = b0 * (e**(b1 * t)) o ln(Y) = ln(b0) + (b1 * t).

Stima di curve: Salva

Salva Variabili. Per ogni modello selezionato è possibile salvare i valori previsti, i residui (valore osservato della variabile dipendente meno il valore previsto del modello) e gli intervalli di previsione (limiti superiore e inferiore). I nomi delle nuove variabili e le etichette descrittive vengono visualizzati in una tabella nell'output.

Periodo di previsione. Se come variabile indipendente si seleziona **Tempo** anzichè una variabile nel dataset attivo, è possibile specificare un periodo di previsione al termine delle serie storiche. È possibile scegliere una delle seguenti opzioni:

- **Prevedi dal periodo di stima fino all'ultimo caso.** Consente di prevedere i valori per tutti i casi del file in base ai casi inclusi nel periodo di stima. Il periodo di stima, visualizzato in fondo alla finestra di dialogo, viene definito nella sottofinestra di dialogo dell'opzione Seleziona casi del menu Dati. Se non è stato definito alcun periodo di stima, i valori verranno previsti in base a tutti i casi.
- **Prevedi fino a.** Consente di prevedere i valori fino alla data, all'ora o al numero di osservazione specificato in base ai casi inclusi nel periodo di stima. Questa funzione può essere usata per prevedere valori futuri nelle serie storiche. Le variabili data definite specificano quali caselle di testo è possibile usare per specificare il termine di un periodo di previsione. Se non vengono definite variabili di data, è possibile specificare l'ultimo numero di osservazione (caso).

Per creare variabili di data, utilizzare l'opzione Definisci date, disponibile nel menu Dati.

Regressione dei minimi quadrati parziali

La procedura Regressione parziale dei minimi quadrati consente di stimare i modelli di regressione parziale dei minimi quadrati (PLS, nota anche come "proiezione della struttura latente"). PLS è una tecnica predittiva che rappresenta un'alternativa alla regressione dei minimi quadrati ordinari (OLS), alla correlazione canonica o alla modellazione di equazioni strutturali e si rivela particolarmente utile quando le variabili predittore sono strettamente correlate o quando il numero dei predittori supera il numero dei casi.

PLS combina le funzioni dell'analisi dei componenti principali e della regressione multipla. Consente innanzitutto di estrarre un insieme di fattori latenti che forniscono la maggiore quantità di informazioni possibile sulla covarianza tra le variabili dipendenti e indipendenti. Quindi, una fase di regressione consente di prevedere i valori delle variabili dipendenti mediante la decomposizione delle variabili indipendenti.

Tabelle
La proporzione della varianza spiegata (per fattore latente), i pesi fattoriali latenti, i caricamenti fattori latenti, l'importanza della variabile indipendente nella proiezione (VIP) e le stime del parametro di regressione (per variabile dipendente) vengono tutti generati per impostazione predefinita.

Grafici
VIP (Variable importance in projection), punteggi fattore, pesi fattore per i primi tre fattori latenti e distanza per il modello sono tutti prodotti dalla scheda Opzioni.

Considerazioni sui dati

Livello di misurazione
Le variabili dipendenti e indipendenti (predittore) possono essere di scala, nominali o ordinali. La procedura presume che a tutte le variabili sia stato assegnato il livello di misurazione appropriato;

sebbene sia possibile modificare temporaneamente il livello di misurazione per una variabile facendo clic con il tasto destro del mouse sulla variabile nell'elenco di variabili di origine e selezionando un livello di misurazione dal menu a comparsa. Le variabili categoriali (nominali o ordinali) vengono trattate in maniera equivalente dalla procedura.

Codifica variabile categoriale
La procedura ricodifica temporaneamente le variabili dipendenti categoriali utilizzando le codifiche one-of-c*c* per la durata della procedura. Se ci sono categorie *c* di una variabile, la variabile viene memorizzata come vettori *c* , con la prima categoria indicata (1,0, ..., 0), la categoria successiva (0,1,0, ..., 0), ..., e la categoria finale (0,0, ..., 0, 1). Le variabili dipendenti categoriali vengono rappresentate utilizzando una codifica dummy; ovvero, semplicemente omettendo l'indicatore corrispondente alla categoria di riferimento.

pesi delle frequenze
I valori dei pesi, prima di essere utilizzati vengono arrotondati ai numeri interi più vicini. I casi con pesi mancanti o con pesi inferiori a 0,5 non vengono utilizzati nelle analisi.

Valori mancanti
I valori mancanti di sistema e definiti dall'utente vengono considerati come non validi.

Ridimensionamento
Tutte le variabili di modello sono centrate e standardizzate, comprese le variabili indicatore che rappresentano le variabili categoriali.

Come ottenere la regressione parziale dei minimi quadrati

Dai menu, scegliere:

Analisi > **Regressione** > **Minimi quadrati parziali ...**

1. Selezionare almeno una variabile dipendente.
2. Selezionare almeno una variabile indipendente.

Facoltativamente, è possibile:

- Specificare una categoria di riferimento per le variabili dipendenti categoriali (nominali o ordinali).
- Specificare una variabile da utilizzare come identificativo univoco per l'output casewise e i dataset salvati.
- Specificare un limite superiore per il numero dei fattori latenti da estrarre.

Modello

Specifica effetti del modello. Un modello di effetti principali include tutti gli effetti principali di covariate e fattori. Selezionare **Personalizzato** per specificare le interazioni. È necessario indicare tutti i termini da includere nel modello.

Fattori e covariate. I fattori e le covariate sono elencati.

Modello. Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato **Personalizzato**, è possibile selezionare gli effetti principali e le interazioni desiderate per l'analisi da eseguire.

Crea termini

Per i fattori e le covariate selezionati:

Interazione. Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

Effetti principali. Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti - 2 vie Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti - 3 vie Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti - 4 vie Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti - 5 vie Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Opzioni

La scheda Opzioni consente all'utente di salvare e rappresentare graficamente le stime dei modelli per singoli casi, fattori latenti e predittori.

Per ogni tipo di dati, specificare il nome di un dataset. I nomi dei dataset devono essere univoci. Se si specifica il nome di un dataset esistente, i suoi contenuti vengono sostituiti; altrimenti, viene creato un nuovo dataset.

- **Salva stime per singoli casi.** Salva le seguenti stime del modello casewise: valori previsti, residui, distanza al modello fattore latente e punteggi fattore latente. Inoltre, traccia i punteggi fattoriali latenti.
- **Salva stime per fattori latenti.** Consente di salvare i caricamenti e i pesi fattoriali latenti. Inoltre, traccia i punteggi fattoriali latenti.
- **Salva stime per variabili indipendenti.** Consente di salvare le stime del parametro di regressione e l'importanza delle variabili per la proiezione (VIP). Inoltre, traccia i VIP per fattore latente.

Analisi del vicino più vicino

L'analisi del vicino più vicino è un metodo per la classificazione dei casi basato sulla similarità ad altri casi. Nell'apprendimento automatico, questo metodo è stato sviluppato per riconoscere modelli di dati senza richiedere una corrispondenza esatta con eventuali modelli o casi archiviati. I casi simili sono vicini gli uni agli altri, mentre i casi dissimili sono distanti gli uni dagli altri. Pertanto, la distanza tra due casi rappresenta una misura della loro dissimilarità.

I casi vicini tra loro vengono detti “vicini.” Quando viene presentato un nuovo caso (holdout), viene calcolata la sua distanza da ogni caso del modello. Le classificazioni dei casi più simili, i vicini più vicini, vengono conteggiati e il nuovo caso viene inserito nella categoria contenente il numero maggiore di vicini più vicini.

È possibile specificare il numero di vicini più vicini da esaminare; questo valore viene denominato *k*.

L'analisi del vicino più vicino può essere utilizzata anche per calcolare i valori per un obiettivo continuo. In questa situazione, il valore obiettivo medio o mediano dei vicini più vicini viene utilizzato per ottenere il valore previsto per il nuovo caso.

Considerazioni sui dati dell'analisi del vicino più vicino

Obiettivo e funzioni. L'obiettivo e le funzioni possono essere:

- *Nominale*. Una variabile può essere considerata nominale quando i suoi valori rappresentano categorie prive di classificazione intrinseca (ad esempio, il dipartimento della società in cui lavora un dipendente). Degli esempi di variabili nominali includono la regione, il codice postale e l'affiliazione religiosa.
- *Ordinale*. Una variabile può essere trattata come ordinale quando i suoi valori rappresentano categorie con una classificazione intrinseca (ad esempio, livelli di soddisfazione del servizio da molto insoddisfatti a molto soddisfatti). Degli esempi di variabili ordinali includono i punteggi di atteggiamento che rappresentano i gradi di soddisfazione o fiducia e i punteggi di classificazione delle preferenze.
- *Scala*. Una variabile può essere trattata come una scala (continua) quando i suoi valori rappresentano categorie ordinate con una metrica significativa, in modo che i confronti di distanza tra i valori siano appropriati. Esempi di variabili di scala sono l'età espressa in anni o il reddito espresso in migliaia di Euro.

  Le variabili nominali e ordinali vengono trattate in modo analogo dall'analisi del vicino più vicino. La procedura presume che a ogni variabile sia stato assegnato il livello di misurazione appropriato; tuttavia, è possibile modificare temporaneamente il livello di misurazione per una variabile facendo clic con il tasto destro del mouse sulla variabile nell'elenco di variabili di origine e selezionando un livello di misurazione dal menu a comparsa.

L'icona accanto a ciascuna variabile nell'elenco delle variabili identifica il livello di misurazione e il tipo di dati.

Tabella 3. Icone del livello di misurazione				
	Numerico	**Stringa**	**Dati**	**Ora**
Scala (continuo)		n/d		
Ordinale				
Nominale				

Codifiche variabili categoriali. La procedura ricodifica temporaneamente le variabili dipendenti e indipendenti categoriali utilizzando le codifiche one-of-*c* per la durata della procedura. Se ci sono le categorie *c* di una variabile, allora la variabile viene memorizzata come vettori c, con la prima categoria denotata (1,0, ..., 0), la categoria successiva (0,1,0, ..., 0), ..., e la categoria finale (0,0, ..., 0, 1).

Questo schema di codifica aumenta la dimensionalità dello spazio delle funzioni. In particolare, il numero totale di dimensioni è pari al numero di predittori di scala più il numero di categorie in tutti i predittori categoriali. Ne consegue che questo schema di codifica può generare un addestramento più lento. Se l'addestramento del vicino più vicino sta procedendo molto lentamente, è possibile cercare di ridurre il numero di categorie nei predittori categoriali mediante la combinazione di categorie simili o casi di rilascio con categorie estremamente rare prima dell'esecuzione della procedura.

Tutte le codifiche one-of-*c* si basano sui dati di addestramento, anche se è definito un campione holdout (consultare “Partizioni ” a pagina 207). Pertanto, se il campione holdout contiene casi con categorie di predittori assenti nei dati di addestramento, non è possibile calcolarne il punteggio. Se il campione holdout contiene casi con categorie di variabili dipendenti assenti nei dati di addestramento, è invece possibile calcolarne il punteggio.

Modifica della scala. Le funzioni di scala vengono normalizzate per impostazione predefinita. Tutta la modifica della scala viene eseguita in base ai dati di addestramento, anche se è definito un campione holdout (consultare “Partizioni ” a pagina 207). Se si specifica una variabile per definire le partizioni, è importante che tali funzioni abbiano distribuzioni simili nei campioni di addestramento e holdout. Ad esempio, utilizzare la procedura Esplora per esaminare le distribuzioni nelle partizioni.

Peso della frequenza. I pesi della frequenza vengono ignorati da questa procedura.

Replica risultati. La procedura utilizza la generazione di numeri casuali durante l'assegnazione casuale delle partizioni e delle occorrenze con convalida incrociata. Se si desidera replicare in modo esatto i risultati, oltre ad utilizzare le impostazioni della stessa procedura, impostare un seme per Mersenne Twister (consultare “Partizioni ” a pagina 207) oppure utilizzare le variabili per definire le partizioni e le occorrenze con convalida incrociata.

Per ottenere un'analisi del vicino più vicino

Dai menu, scegliere:

Analizzare > **Classifica** > **Vicina più vicina ...**

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

1. Specificare una o più funzioni, che possono essere considerate variabili o predittori indipendenti in presenza di un obiettivo.

   Obiettivo (facoltativo). Se non è specificato un obiettivo (risposta o variabile dipendente), la procedura trova solo i *k* vicini più vicini, non viene eseguita alcuna classificazione o previsione.

   Normalizza funzioni di scala. Le funzioni normalizzate hanno lo stesso intervallo di valori, il che può migliorare la prestazione dell'algoritmo di stima. Viene utilizzata la normalizzazione adattata, [2*(x–min)/(max–min)]–1. I valori normalizzati adattati sono compresi tra –1 e 1.

Identificativo dei casi focali (facoltativo). Consente di contrassegnare casi di particolare interesse. Ad esempio, un ricercatore desidera determinare se i punteggi del test da un distretto scolastico, il caso focale, possono essere confrontati con quelli di distretti scolastici simili. Ricorrerà all'analisi del vicino più vicino per individuare i distretti scolastici con le maggiori analogie in termini di uno specifico insieme di funzioni. Procederà quindi al confronto tra i punteggi del test del distretto scolastico focale e quelli dei vicini più vicini.

I casi focali possono essere usati anche negli studi clinici per selezionare casi di controllo simili a casi clinici. I casi focali vengono visualizzati nella tabella dei *k* vicini più vicini e delle distanze, nel grafico dello spazio delle funzioni, in quello di peer e nella mappa dei quadranti. Le informazioni sui casi focali vengono salvate nei file specificati nella scheda Output.

I casi con un valore positivo per la variabile specificata vengono trattati come casi focali. Non è consentito specificare una variabile senza valori positivi.

Etichetta caso (facoltativo). Ai casi vengono applicate etichette utilizzando questi valori nel grafico dello spazio delle funzioni, in quello dei peer e nella mappa dei quadranti.

Campi con livello di misurazione sconosciuto

L'avviso Livello di misurazione viene visualizzato quando il livello di misurazione di una o più variabili (campi) del dataset è sconosciuto. Poiché influisce sul calcolo dei risultati di questa procedura, il livello di misurazione deve essere definito per tutte le variabili.

Esegui scansione dati. Legge i dati del dataset attivo e assegna un livello di misurazione predefinito a tutti i campi con livello di misurazione sconosciuto. Con dataset di grandi dimensioni, questa operazione può richiedere del tempo.

Assegna manualmente. Apre una finestra di dialogo che elenca tutti i campi con livello di misurazione sconosciuto, mediante la quale è possibile assegnare un livello di misurazione a questi campi. Il livello di misurazione si può assegnare anche nella Vista variabile dell'Editor dei dati.

Dal momento che il livello di misurazione è importante per questa procedura, è possibile accedere alla finestra di dialogo per la sua esecuzione solo quando per tutti i campi è stato definito un livello di misurazione.

Vicini

Numero di Vicini più vicini (k). Specificare il numero di vicini più vicini. L'utilizzo di un numero maggiore di vicini non garantisce necessariamente un modello più preciso.

Se nella tabella Variabili è specificato un obiettivo, in alternativa è possibile indicare un intervallo di valori e lasciare che sia la procedura a scegliere il "miglior" numero di vicini all'interno di tale intervallo. Il metodo utilizzato per stabilire il numero di vicini più vicini dipende dalla necessità o meno della selezione delle funzioni nella scheda Funzioni.

- Se la selezione delle funzioni è attiva, viene eseguita per ciascun valore di *k* nell'intervallo richiesto e viene selezionato il *k* (con il relativo insieme di funzioni) con il tasso di errore più basso (o l'errore più basso della somma dei quadrati se l'obiettivo è una scala).
- Se la selezione delle funzioni non è attiva, viene utilizzata la convalida incrociata con occorrenze *V* per selezionare il “miglior” numero di vicini. Per informazioni sul controllo dell'assegnazione delle occorrenze, vedere la scheda Partizione.

Calcolo delle distanze. Metrica utilizzata per specificare la metrica di distanza per la misurazione della similarità dei casi.

- **Metrica euclidea.** La distanza tra due casi, x e y, è pari alla radice quadrata della somma, in tutte le dimensioni, delle differenze al quadrato tra i valori di tali casi.
- **Metrica City Block.** La distanza tra due casi è pari alla somma, in tutte le dimensioni, delle differenze assolute tra i valori di tali casi. È denominata anche "distanza di Manhattan".

Se si desidera, qualora nella scheda Variabili sia specificato un obiettivo, è possibile scegliere di pesare le funzioni in base alla loro importanza normalizzata durante il calcolo delle distanze. L'importanza della

funzione di un predittore si calcola dividendo il tasso di errore o l'errore della somma dei quadrati relativo al modello senza il predittore per il tasso di errore o l'errore della somma dei quadrati relativo al modello completo. L'importanza normalizzata si calcola ripesando i valori di importanza della funzione in modo che la somma sia pari a 1.

Previsioni per obiettivi di scala. Se nella scheda Variabili è specificato un obiettivo di scala, questo indica se il valore previsto è calcolato in base alla media o al valore mediano dei vicini più vicini.

Funzioni

La scheda Funzioni consente di richiedere e specificare opzioni per la selezione delle funzioni quando nella scheda Variabili è specificato un obiettivo. Per impostazione predefinita, per la selezione delle funzioni vengono prese in considerazione tutte le funzioni, ma è possibile selezionare un sottoinsieme di funzioni da forzare nel modello.

Criteri di arresto. A ogni fase, viene presa in considerazione, per essere inclusa nell'insieme dei modelli, la funzione il cui inserimento nel modello determina l'errore minore (calcolato come tasso di errore per gli obiettivi categoriali e come errore della somma dei quadrati per gli obiettivi di scala). La selezione in avanti prosegue fino al raggiungimento della condizione specificata.

- **Numero specificato di funzioni.** L'algoritmo inserisce un numero fisso di funzioni oltre a quelle forzate nel modello. Specificare un intero positivo. La riduzione dei valori del numero da selezionare dà origine a un modello più parsimonioso, con il rischio di perdere funzioni importanti. L'aumento dei valori del numero da selezionare consente di acquisire tutte le funzioni importanti, con il rischio però di aggiungere funzioni che finiscono per moltiplicare l'errore del modello.
- **Modifica minima nel rapporto di errore assoluto.** L'algoritmo si arresta quando la variazione del rapporto di errore assoluto indica che il modello non può essere migliorato ulteriormente aggiungendo altre funzioni. Specificare un numero positivo. La diminuzione dei valori della modifica tenderà ad includere altre funzioni, con il rischio di includere funzioni che non aggiungono molto valore al modello. L'aumento del valore della variazione minima, invece, tende a impedire l'inserimento di altre funzioni, con il rischio di perderne alcune importanti per il modello. Il valore “ottimale” della modifica minima dipenderà dai dati e dall'applicazione. Per assistenza nella valutazione delle funzioni più importanti, vedere il log degli errori relativi alla selezione delle funzioni nell'output. Per ulteriori informazioni, consultare l'argomento “Log degli errori relativi alla selezione delle funzioni ” a pagina 211.

Partizioni

La scheda Partizioni consente di suddividere il dataset in sottoinsiemi di addestramento e holdout e, se possibile, di assegnare casi a occorrenze con convalida incrociata.

Partizioni di addestramento e holdout. Questo gruppo specifica il metodo di partizionamento del dataset attivo in campioni di addestramento e holdout. Il **campione di addestramento** include i record di dati utilizzati per formare il modello di vicino più vicino; una percentuale di casi nel dataset deve essere assegnata al campione di addestramento per ottenere un modello. Il **campione holdout** è un insieme indipendente di record di dati utilizzato per valutare il modello finale; l'errore per il campione holdout fornisce una stima "attendibile" della capacità predittiva del modello poiché i casi di holdout non sono stati utilizzati per generare il modello.

- **Assegna in modo casuale i casi alle partizioni.** Specificare la percentuale di casi da assegnare al campione di addestramento. Il resto viene assegnato al campione holdout.
- **Usa variabile per assegnare casi.** Specificare una variabile numerica che assegni ogni caso nel dataset attivo al campione di addestramento o holdout. I casi con valore positivo nella variabile vengono assegnati al campione di addestramento, quelli con valore pari a 0 o negativo al campione holdout. I casi con un valore mancante di sistema vengono esclusi dall'analisi. I valori mancanti definiti dall'utente per la variabile partizione sono sempre considerati validi.

Occorrenze con convalida incrociata. La convalida incrociata con occorrenze *V* viene utilizzata per determinare il "miglior" numero di vicini. Per motivi legati alla prestazione, la convalida incrociata non è disponibile se si utilizza la selezione delle funzioni.

La convalida incrociata suddivide il campione in una serie di sottocampioni o occorrenze. I modelli di vicino più vicino vengono quindi generati escludendo di volta in volta i dati da ciascun sottocampione. Il primo modello si basa su tutti i casi eccetto quelli contenuti nella prima occorrenza del campione, il secondo modello si basa su tutti i casi eccetto quelli contenuti nella seconda occorrenza del campione e così via. Il rischio di errore per ciascun modello viene stimato applicando il modello al sottocampione escluso al momento della generazione del modello stesso. Il "miglior" numero di vicini più vicini è quello che genera l'errore più basso in tutte le occorrenze.

- **Assegna in modo casuale i casi alle occorrenze.** Specificare il numero di occorrenze da utilizzare per la convalida incrociata. I casi vengono assegnati in modo casuale alle occorrenze, numerati da 1 a *V*, il numero di occorrenze.
- **Usa variabile per assegnare casi.** Specificare una variabile numerica che assegni ogni caso nel dataset attivo a un'occorrenze. La variabile deve essere numerica e assumere valori da 1 a *V*. Se mancano valori in questo intervallo e su qualsiasi suddivisione se i file suddivisi sono in vigore, questo causi un errore.

Imposta seme per Mersenne Twister. Impostando un seme è possibile replicare le analisi. L'utilizzo di questo controllo è simile all'impostazione di Mersenne Twister come generatore attivo e alla specifica di un punto iniziale fisso nella finestra di dialogo Generatori di numeri casuali, con la differenza importante che impostando il seme in questa finestra di dialogo preserverà lo stato corrente del generatore di numeri casuali e ripristinerà tale stato una volta completata l'analisi.

Salva

Nomi delle variabili salvate. La generazione automatica del nome assicura il mantenimento di tutto il lavoro. I nomi personalizzati consentono di eliminare/sostituire i risultati di precedenti esecuzioni senza dover prima eliminare le variabili salvate nell'Editor dei dati.

Variabili da salvare

- **Valore o categoria prevista.** Viene salvato il valore previsto per un obiettivo di scala o la categoria prevista per un obiettivo categoriale.
- **Probabilità prevista.** Vengono salvate le probabilità previste per un obiettivo categoriale. Una variabile separata viene salvata per ognuna delle prime *n* categorie, dove *n* viene specificato nel comando **Numero massimo di categorie da salvare per l'obiettivo categoriale**.
- **Variabile partizione di addestramento/holdout.** Se i casi vengono assegnati in modo casuale ai campioni addestramento e holdout nella scheda Partizioni, viene salvato il valore della partizione (di addestramento o holdout) a cui il caso è stato assegnato.
- **Variabile occorrenza con convalida incrociata.** Se nella scheda Partizioni alle occorrenze con convalida incrociata vengono assegnati casi in modo casuale, viene salvato il valore dell'occorrenza a cui è stato assegnato il caso.

Output

Output visualizzatore

- **Riepilogo elaborazione casi.** Visualizza la tabella di riepilogo di elaborazione dei casi, che riepiloga il numero di casi inclusi ed esclusi dall'analisi, in totale e per campioni di addestramento e holdout.
- **Grafici e tabelle.** Visualizza l'output relativo al modello, tra cui tabelle e grafici. Le tabelle nella vista modello comprendono *k* vicini più vicini e le distanze per i casi focali, la classificazione delle variabili di risposta categoriali e un riepilogo degli errori. L'output grafico nella vista modello include un log degli errori relativi alla selezione, il grafico dell'importanza delle funzioni, quello dello spazio delle funzioni e dei peer e la mappa dei quadranti. Per ulteriori informazioni, consultare l'argomento “Vista Modello ” a pagina 209.

File

- **Esporta modello in file XML.** È possibile utilizzare questo file modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio. Questa opzione non è disponibile se sono stati definiti file suddivisi.

- **Esporta distanze tra casi focali e k vicini più vicini.** Per ogni caso focale viene creata una variabile distinta per ciascun *k* vicino più vicino (dal campione di addestramento) del caso focale stesso e delle corrispondenti *k* distanze più vicine.

Opzioni

Valori mancanti definiti dall'utente. Le variabili categoriali devono contenere valori validi per un caso per essere incluse nell'analisi. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente devono essere considerati validi nell'ambito delle variabili categoriali.

I valori mancanti di sistema e i valori mancanti relativi alle variabili di scala vengono sempre considerati non validi.

Vista Modello

Selezionando **Grafici e tabelle** nella scheda Output, nel Visualizzatore viene creato un oggetto Modello vicino più vicino. Attivando l'oggetto con un doppio clic, si accede a una vista interattiva del modello La vista modello ha una finestra a due pannelli:

- Nel primo è presente una panoramica del modello denominata "vista principale".
- Nel secondo, invece, possono essere visualizzate due tipologie di vista:

  La vista modello ausiliaria mostra ulteriori informazioni sul modello, pur non concentrandosi su quest'ultimo.

  La vista collegata mostra invece i dettagli relativi a una funzione del modello quando l'utente esegue il drill-down di parte della vista principale.

Per impostazione predefinita, nel primo pannello viene visualizzato lo spazio di funzioni e nel secondo il grafico dell'importanza delle variabili. Se quest'ultimo grafico non è disponibile (se, cioè, nella scheda Funzioni non è stato selezionato **Pesa funzioni in base all'importanza**), viene visualizzata la prima vista presente nell'elenco a discesa Vista.

Quando per una vista non sono disponibili informazioni, la voce corrispondente nell'elenco a discesa Vista viene disattivata.

Spazio di funzioni

Il grafico dello spazio delle funzioni è un grafico interattivo relativo allo spazio delle funzioni (o al sottospazio, se sono presenti più di tre funzioni). Ogni asse rappresenta una funzione nel modello e l'ubicazione dei punti nel grafico indica i valori di tali funzioni per i casi nelle partizioni di addestramento e holdout.

Chiavi. Oltre a rappresentare i valori delle funzioni, i punti forniscono altre informazioni.

- La forma indica la partizione (Addestramento o Holdout) di cui fa parte un punto.
- Il colore/l'ombreggiatura di un punto indica il valore dell'obiettivo del caso (i diversi valori di colore corrispondono alle categorie di un obiettivo categoriale, mentre le ombreggiature indicano l'intervallo di valori di un obiettivo continuo). Il valore indicato per la partizione di addestramento è quello osservato, mentre per la partizione di holdout è indicato quello previsto. Se non viene specificato alcun obiettivo, questo simbolo non viene visualizzato.
- Schemi più marcati indicano che un caso è focale. I casi focali vengono visualizzati collegati ai relativi *k* vicini più vicini.

Controlli e interattività. Nel grafico è disponibile una serie di comandi per esplorare lo spazio di funzioni.

- È possibile scegliere il sottoinsieme di funzioni da visualizzare nel grafico e cambiare le funzioni da rappresentare nelle dimensioni.
- I “casi focali” sono semplicemente i punti selezionati nel grafico Spazio di funzioni. Se è stata specificata una variabile per casi focali, inizialmente verranno selezionati i punti che rappresentano i casi focali. Qualsiasi punto può comunque diventare temporaneamente un caso focale se viene

selezionato. Vengono utilizzati i controlli “normali” per la selezione del punto; facendo clic su un punto, questo viene selezionato e vengono deselezionati tutti gli altri; facendo clic su un punto tenendo premuto il tasto Ctrl, lo si aggiunge all'insieme di punti selezionati. Le viste collegate, ad esempio il grafico dei peer, vengono automaticamente aggiornate in base ai casi selezionati nello spazio di funzioni.

- È possibile modificare il numero di vicini più vicini (*k*) da visualizzare per i casi focali.
- Passando il mouse sopra un punto del grafico, viene visualizzato un suggerimento con il valore dell'etichetta del caso (o il numero del caso se non sono state definite etichette), oltre ai valori osservati e previsti dell'obiettivo.
- Il pulsante “Reimposta” consente di ripristinare lo stato originale dello Spazio di funzioni.

Aggiunta e rimozione di campi/variabili

È possibile aggiungere nuovi campi/nuove variabili allo spazio di funzioni o rimuovere quelli attualmente visualizzati.

Palette Variabili

La palette Variabili deve essere visualizzata prima di poter aggiungere e rimuovere le variabili. Per visualizzare la palette Variabili, il Visualizzatore modelli deve essere in modalità Modifica e un caso deve essere selezionato nello spazio di funzioni.

1. Per impostare il Visualizzatore modelli in modalità Modifica, dal menu scegliere:

   Visualizza > **Modifica modo**
2. Una volta in Modalità Modifica, fare clic su un caso nello spazio di funzioni.
3. Per visualizzare la palette Variabili, dai menu scegliere:

   Visualizza > **Le tavole** > **Variabili**

   La palette Variabili elenca tutte le variabili nello spazio di funzioni. L'icona accanto al nome della variabile indica il livello di misurazione della variabile.
4. Per cambiare temporaneamente il livello di misurazione di una variabile, fare clic con il tasto destro del mouse sulla variabile nella palette Variabili e scegliere un'opzione.

Zone delle variabili

Le variabili vengono aggiunte alle "zone" nello spazio di funzioni. Per visualizzare le zone, iniziare a trascinare una variabile dalla palette Variabili oppure selezionare **Mostra zone**.

Lo spazio di funzioni ha delle zone per gli assi *x*, *y* e *z*.

Spostamento delle variabili nelle zone

Di seguito vengono fornite alcune regole generali e suggerimenti per lo spostamento delle variabili nelle zone.

- Per spostare una variabile in una zona, fare clic e trascinare la variabile dalla palette Variabili e rilasciarla nella zona. Se si sceglie **Mostra zone**, è anche possibile fare clic con il tasto destro del mouse su una zona e selezionare una variabile che si desidera aggiungere alla zona.
- Se si trascina una variabile dalla palette Variabili in una zona già occupata da un'altra variabile, la vecchia variabile viene sostituita con quella nuova
- Se si trascina una variabile da una zona in una zona già occupata da un'altra variabile, le variabili si scambiano posizione.
- Facendo clic su X in una zona, la variabile viene rimossa dalla zona in questione.
- Se vi sono più elementi grafici nella visualizzazione, ciascun elemento grafico può avere le proprie zone delle variabili associate. Selezionare prima l'elemento grafico.

Importanza della variabile

Di solito è opportuno concentrare la modellazione sulle variabili più rilevanti, lasciando perdere o ignorando le meno importanti. In questo senso può essere utile il grafico dell'importanza delle variabili, che indica l'importanza relativa di ciascuna variabile nella stima del modello. Dal momento che i valori sono relativi, la somma dei valori di tutte le variabili visualizzate è pari a 1,0. L'importanza delle variabili non ha nulla a che vedere con la precisione del modello. Riguarda unicamente l'importanza di ciascuna variabile per l'elaborazione di una previsione, non il grado di precisione di quest'ultima.

Equivalenti

In questo grafico vengono visualizzati i casi focali e i relativi *k* vicini più vicini per ciascuna funzione e per l'obiettivo. È disponibile se nello spazio di funzioni è selezionato un caso focale.

Collegamenti. Il grafico dei peer è collegato allo Spazio di funzioni in due modi.

- I casi selezionati (focali) nello spazio di funzioni vengono visualizzati nel grafico dei peer, insieme ai relativi *k* vicini più vicini.
- Il valore di *k* selezionato nello spazio di funzioni viene utilizzato nel grafico dei peer.

Distanze dei vicini più vicini

In questa tabella vengono visualizzati i vicini più vicini e le distanze più vicine *k* solo per i casi focali. È disponibile se nella scheda Variabili è specificato un identificativo dei casi focali e mostra soltanto i casi focali identificati da questa variabile.

Ogni riga della:

- Colonna **Caso focale** contiene il valore della variabile di etichetta relativa al caso focale. Se non sono definite etichette dei casi, la colonna contiene il numero del caso focale.
- *i*ª colonna del gruppo Vicini più vicini contiene il valore della variabile di etichetta dei casi relativa al *i*° vicino più vicino del caso focale. Se non sono definite etichette dei casi, la colonna contiene il numero di caso del *i*° vicino più vicino del caso focale.
- *i*ª colonna del gruppo Distanze più vicine contiene la distanza del *i*° vicino più vicino dal caso focale.

Mappa dei quadranti

Il grafico mostra i casi focali e i relativi *k* vicini più vicini su un grafico a dispersione (o un grafico a punti a seconda del livello di misurazione dell'obiettivo) con l'obiettivo sull'asse *y* e una funzione di scala sull'asse *x*, il tutto suddiviso in pannelli in base alle funzioni. È disponibile se nello spazio di funzioni è presente un obiettivo ed è selezionato un caso focale.

- Per le variabili continue, nella partizione di addestramento in corrispondenza delle medie variabile vengono tracciate righe di riferimento.

Log degli errori relativi alla selezione delle funzioni

I punti presenti nel grafico mostrano l'errore (in termini di tasso di errore o di errore della somma dei quadrati a seconda del livello di misurazione dell'obiettivo) sull'asse *y* del modello, con la funzione elencata sull'asse *x* (inoltre, a sinistra sull'asse *x* sono presenti tutte le funzioni). Il grafico è disponibile se è presente un obiettivo ed è attiva la selezione funzioni.

Log degli errori relativi alla selezione di k

I punti presenti nel grafico mostrano l'errore (in termini di tasso di errore o di errore della somma dei quadrati a seconda del livello di misurazione dell'obiettivo) sull'asse *y* del modello, con il numero di vicini più vicini (*k*) sull'asse *x*. Il grafico è disponibile se è presente un obiettivo ed è attiva la selezione *k*.

Log degli errori relativi alla selezione k e alla selezione delle funzioni

Si tratta di grafici di selezione delle funzioni (vedi “Log degli errori relativi alla selezione delle funzioni ” a pagina 211), paneggiati da *k*. Questo grafico è disponibile se c'è un obiettivo e *k* e la selezione delle funzioni sono entrambi in vigore.

Tabella di classificazione

Nella tabella viene visualizzata la classificazione incrociata dei valori osservati dell'obiettivo rispetto a quelli previsti, suddivisi per partizione. È disponibile se è presente un obiettivo di tipo categoriale.

- La riga **(Mancante)** della partizione di holdout contiene casi di holdout con valori mancanti sull'obiettivo. Questi casi contribuiscono al campione di Holdout: valori di Percentuale globale, ma non ai valori di Percentuale di correttezza.

Riepilogo degli errori

La tabella è disponibile in presenza di una variabile obiettivo. Visualizza l'errore associato al modello; la somma dei quadrati per un obiettivo continuo e il tasso di errore (100% – percentuale globale di correttezza) per un obiettivo categoriale.

Analisi discriminante

L'analisi discriminante crea un modello di previsione per l'appartenenza al gruppo. Il modello è costituito da una funzione discriminante oppure, per più di due gruppi, da un insieme di funzioni discriminanti, in base alle combinazioni lineari delle variabili predittore che forniscono la migliore discriminazione tra i gruppi. Le funzioni vengono generate da un campione di casi per cui è nota l'appartenenza al gruppo; è quindi possibile applicare le funzioni ai nuovi casi con misurazioni per le variabili predittore, ma di cui non è nota l'appartenenza al gruppo.

Nota: la variabile di raggruppamento può avere più di due valori. I codici per la variabile di raggruppamento devono tuttavia essere interi, ed è necessario specificare i valori massimo e minimo corrispondenti. I casi con valore non compreso tra i due estremi specificati vengono esclusi dall'analisi.

Esempio. In media, il consumo calorico giornaliero degli abitanti delle zone temperate è maggiore di quello di chi vive ai tropici. Nelle zone temperate, inoltre, si riscontra una maggiore percentuale di persone che vivono in ambiente urbano. Un ricercatore desidera combinare queste informazioni in una funzione per determinare le modalità di discriminazione tra i due gruppi di paesi. Il ricercatore ritiene opportuno prendere in considerazione anche le dimensioni della popolazione e informazioni di carattere economico. L'analisi discriminante consente di valutare i coefficienti della funzione discriminante lineare, analoga alla parte destra di un'equazione di regressione lineare multipla. In altri termini, utilizzando i coefficienti *a*, *b*, *c* e *d* si ottiene la funzione:

```
D = a * climate + b * urban + c * population + d * gross domestic product per capita
```

Se queste variabili sono utili per la discriminazione tra le due zone climatiche, i valori di *D* per i paesi temperati saranno diversi da quelli relativi ai paesi tropicali. Se è necessario usare un metodo di selezione delle variabili stepwise, nella funzione non si dovranno includere tutte e quattro le variabili.

Statistiche. Per ogni variabile: medie, deviazioni standard, ANOVA univariate. Per ogni analisi: *M* di Box, matrice di correlazione entro gruppi, matrice di covarianza entro gruppi, matrice di covarianza di gruppi separati, matrice di covarianza totale. Per ogni funzione discriminante canonica: autovalore, percentuale di varianza, correlazione canonica, lambda di Wilks, chi-quadrato. Per ogni fase: probabilità a priori, coefficienti di funzione di Fisher, coefficienti di funzione non standardizzati, lambda di Wilks per ogni funzione canonica.

Considerazioni sui dati relative all'analisi discriminante

Dati. La variabile di raggruppamento deve includere un numero limitato di categorie distinte, codificate come interi. Le variabili indipendenti nominali devono essere ricodificate in forma di variabili dummy o di contrasto.

Ipotesi. I casi devono essere indipendenti. Le variabili predittore devono avere una distribuzione normale multivariata e le matrici di varianza-covarianza entro gruppi devono essere uguali in tutti i gruppi. Si assume che le appartenenze ai gruppi si escludano reciprocamente (ovvero che nessun caso appartenga a più gruppi) e che ciascun caso appartenga a un gruppo. La procedura è più efficace se l'appartenenza ai gruppi è una variabile categoriale effettiva; se l'appartenenza ai gruppi si basa sui valori di una variabile continua (ad esempio, QI massimo e QI minimo), è opportuno utilizzare la regressione lineare per avvalersi delle informazioni più dettagliate disponibili nella variabile continua.

Per ottenere un'analisi discriminante

1. Dai menu, scegliere:

   Analizza > **Classifica** > **Discriminante...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una variabile di raggruppamento con valori interi e fare clic su **Definisci intervallo** per specificare le categorie desiderate.
3. Selezionare le variabili indipendenti o le variabili stimatore. (Se la variabile di raggruppamento non include valori interi, utilizzando il comando Ricodifica automatica del menu Trasforma è possibile crearne una che includa tali valori).
4. Selezionare il metodo di inserimento delle variabili indipendenti.
   - **Inserisci indipendenti insieme.** Inserisce contemporaneamente tutte le variabili indipendenti che soddisfano i criteri di tolleranza.
   - **Usa metodo stepwise.** Usa l'analisi stepwise per controllare l'inserimento della variabile e la rimozione.
5. È inoltre possibile selezionare i casi utilizzando una variabile di selezione.

Analisi discriminante: Definisci intervallo

Specificare il valore minimo e massimo della variabile di raggruppamento da utilizzare per l'analisi. I casi con valori che non rientrano in tale intervallo non vengono utilizzati nell'analisi discriminante, ma vengono classificati in uno dei gruppi esistenti in base ai risultati dell'analisi stessa. I valori massimo e minimo devono essere interi.

Analisi discriminante: Seleziona casi

Per selezionare i casi da usare nell'analisi:

1. Scegliere la variabile di selezione nella finestra di dialogo Analisi discriminante.
2. Fare clic su**Valore** per immettere un intero come variabile di selezione.

Le funzioni discriminanti verranno derivate solo in base ai casi che prevedono tale valore per la variabile di selezione. Le statistiche e i risultati della classificazione vengono generati sia per i casi selezionati che per i casi non selezionati. Tale processo consente di classificare i nuovi casi sulla base dei dati preesistenti nonché ripartire i dati in sottoinsiemi di addestramento e test con i quali eseguire la convalida del modello generato.

Analisi discriminante: Statistiche

Descrittive. Le opzioni disponibili sono medie (incluse le deviazioni standard), ANOVA univariate e test *M* di Box.

- *Medie*. Visualizza le medie totali e di gruppo, nonché le deviazioni standard per le variabili indipendenti.
- *ANOVA univariate*. Effettua l'analisi della varianza univariata a una via per verificare l'uguaglianza delle medie di gruppo per ciascuna variabile indipendente.

- *M Box*. Un test per l'uguaglianza delle matrici di covarianza del gruppo. Per dimensioni del campione sufficientemente elevate, un valore P non significativo vuol dire che non ci sono sufficienti prove che le matrici differiscano. Il test è sensibile a scostamenti dalla normalità multivariata.

Coefficienti di funzione. Le opzioni disponibili sono i coefficienti di correlazione di Fisher e i coefficienti non standardizzati.

- *Di Fisher*. Visualizza i coefficienti della funzione di classificazione di Fisher che possono essere utilizzati direttamente per la classificazione. Viene ottenuto un insieme separato di coefficienti di funzioni di classificazione per ciascun gruppo e a ogni caso viene assegnato al gruppo in cui ottiene il più alto punteggio discriminante (valore della funzione di classificazione).
- *Non standardizzati*. Visualizza i coefficienti della funzione discriminante non standardizzata.

Matrici. Le matrici di coefficienti disponibili per le variabili indipendenti sono: matrice di correlazione entro gruppi, matrice di covarianza entro gruppi, matrice di covarianza gruppi separati e matrice di covarianza totale.

- *Correlazione entro gruppi*.Visualizza la matrice di correlazione entro gruppi ottenuta mediando le matrici di covarianza di tutti i gruppi prima di calcolare le correlazioni.
- *Covarianza entro gruppi*.Visualizza una matrice di covarianza entro i gruppi nel pool, che potrebbe variare rispetto alla matrice di covarianza totale. La matrice viene ottenuta calcolando la media delle matrici di covarianza separate per tutti i gruppi.
- *Covarianza di gruppi separati*.Visualizza matrici di covarianza separate per ciascun gruppo.
- *Covarianza totale*.Visualizza una matrice di covarianza di tutti i casi come se provenissero da un unico campione.

Analisi discriminante: Metodo stepwise

Metodo. Selezionare la statistica da utilizzare per inserire o rimuovere nuove variabili. Le alternative disponibili sono: Le alternative disponibili sono la lambda di Wilks, la varianza non spiegata, la distanza Mahalanobis, il più piccolo rapporto *F* e il *V*di Rao. Con il *V*di Rao è possibile specificare l'aumento minimo in *V* per una variabile da inserire.

- *Lambda di Wilks*.Un metodo di selezione delle variabili per l'analisi discriminante stepwise che sceglie le variabili da inserire nell'equazione in base a quanto esse riducono il lambda di Wilks. A ogni fase, viene inserita la variabile che minimizza il Lambda di Wilks globale.
- *Varianza non spiegata*.Ad ogni passo viene inserita la variabile che riduce al minimo la somma della variazione spiegata fra gruppi.
- *Distanza di Mahalanobis*.Una misura di quanto i valori di un caso sulle variabili indipendenti differiscono dalla media di tutti i casi. Un'elevata distanza di Mahalanobis identifica un caso come caratterizzato da valori estremi per una o più variabili indipendenti.
- *Rapporto F più piccolo*.Un metodo di selezione delle variabili nelle analisi per passi basato sulla massimizzazione di un rapporto F valutato tramite la distanza di Mahalanobis tra gruppi.
- *V di Rao*. Una misura delle differenze tra le medie di gruppo. Detta anche traccia di Lawley-Hotelling. Ad ogni fase viene inserita la variabile che massimizza l'aumento della V di Rao. Dopo aver selezionato questa opzione, immettere il valore minimo che una variabile deve avere per essere inserita nell'analisi.

Criteri. Le alternative disponibili sono **Utilizza valore F** e **Utilizza probabilità di F**. Inserire i valori per l'inserimento e la rimozione delle variabili.

- *Usa valore di F*.Una variabile viene immessa nel modello se il suo valore F è maggiore del valore di immissione e viene rimossa se il valore F è minore del valore di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi e Inserimento deve essere maggiore di Rimozione Abbassando il valore di inserimento e/o alzando quello di rimozione Per rimuovere un numero maggiore di variabili dal modello, aumentare il valore di rimozione.
- *La probabilità di utilizzo di F*. Una variabile viene immessa nel modello se il livello di significatività del suo valore F è minore del valore di inserimento e viene rimossa se il livello di significatività è maggiore del valore di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi

e Inserimento deve essere minore di Rimozione. Alzando il valore di inserimento e/o abbassando quello di rimozione Per rimuovere un numero maggiore di variabili dal modello, ridurre il valore di rimozione.

Visualizza. Riepilogo di passi visualizza le statistiche relative a tutte le variabili dopo ogni passaggio; **F per distanze pairwise** visualizza una matrice di rapporti *F* pairwise per ogni coppia di gruppi.

Analisi discriminante: Classificazione

Probabilità a priori. Questa opzione determina se i coefficienti di classificazione vengono adattati per una conoscenza a priori dell'appartenenza al gruppo.

- **Tutti i gruppi uguali.** Si presuppongono probabilità a priori uguali per tutti i gruppi, senza effetti sui coefficienti.
- **Calcola dalle dimensioni dei gruppi.** Le dimensioni del gruppo osservate determinano le probabilità a priori dell'appartenenza al gruppo. Ad esempio, se il 50% delle osservazioni incluse nell'analisi rientrano nel primo gruppo, il 25% nel secondo e il 25% nel terzo, i coefficienti di classificazione vengono adattati in modo da aumentare la probabilità di appartenenza nel primo gruppo rispetto agli altri due.

Visualizza. Le opzioni di visualizzazione disponibili sono: risultati casewise, tabella riassuntiva e classificazione leave-one-out.

- *Risultati casewise*.Visualizza per ciascun caso i codici del gruppo effettivo, del gruppo previsto, della probabilità a posteriori e del punteggio discriminante.
- *Tabella Riepilogo*.Il numero di casi assegnati correttamente o erroneamente a ogni gruppo in base all'analisi discriminante. A volte detta "Matrice confusione".
- *Classificazione leave-one-out*.Ogni caso nell'analisi viene classificato tramite le funzioni ricavate da tutti i casi diversi da tale caso. È noto anche come "metodo U".

Sostituisci valori mancanti con la media. Selezionare questa opzione per sostituire un valore mancante con la media di una variabile indipendente, solo durante la fase di classificazione.

Usa matrice di covarianza. È possibile classificare i casi utilizzando una matrice di covarianza entro gruppi o una matrice di covarianza gruppi separati.

- *Entro gruppi*.La matrice di covarianza entro i gruppi viene utilizzato per classificare casi.
- *Gruppi separati*.Le matrici di covarianza di gruppi separati vengono utilizzate per la classificazione. Poiché la classificazione è basata sulle funzioni discriminanti (non sulle variabili originali), questa opzione non è sempre equivalente alla discriminazione quadratica.

Grafici. Le opzioni disponibili per i grafici sono: gruppi accorpati, gruppi separati e mappa territoriale.

- *Gruppi combinati*.Crea un grafico a dispersione per tutti i gruppi dei primi due valori di funzioni discriminanti. Se c'è una sola funzione, viene invece visualizzato un istogramma.
- *Gruppi separati*.Crea grafici a dispersione di gruppi separati dei primi due valori di funzioni discriminanti. Se c'è una sola funzione, vengono invece visualizzati degli istogrammi.
- *Mappa territoriale*.Un grafico dei confini utilizzati per classificare casi in gruppi in base ai valori di una funzione. I numeri corrispondono ai gruppi nei quali vengono classificati i casi. La media per ciascun gruppo è indicata da un asterisco all'interno dei suoi limiti. La mappa non viene visualizzata se c'è una sola funzione discriminante.

Analisi discriminante: Salva

È possibile aggiungere nuove variabili al file di dati attivo. Le opzioni disponibili sono: appartenenza al gruppo previsto (una sola variabile), punteggi discriminanti (una variabile per ciascuna funzione discriminante nella soluzione) e probabilità di appartenenza al gruppo in base ai punteggi discriminanti (una variabile per ciascun gruppo).

È anche possibile esportare le informazioni sul modello nel file specificato in formato XML. È possibile utilizzare questo file modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio.

Funzioni aggiuntive del comando DISCRIMINANT

Il linguaggio della sintassi dei comandi consente inoltre di:

- Eseguire più analisi discriminanti (con un comando) e controllare l'ordine in cui le variabili vengono inserite (con il sottocomando ANALYSIS).
- Specificare le probabilità a priori per la classificazione (con il sottocomando PRIORS).
- Visualizzare le matrici dei modelli e delle strutture ruotate (con il sottocomando ROTATE).
- Limitare il numero di funzioni discriminanti estratte (con il sottocomando FUNCTIONS).
- Limitare la classificazione ai casi selezionati (o non selezionati) per l'analisi (con il sottocomando SELECT).
- Leggere e analizzare la matrice di correlazione (con il sottocomando MATRIX).
- Scrivere una matrice di correlazione da analizzare in seguito (con il sottocomando MATRIX).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Analisi fattoriale

L'analisi fattoriale si propone di identificare le variabili sottostanti, o **fattori**, che illustrano il modello per le correlazioni all'interno di un insieme di variabili osservate. L'analisi fattoriale viene in genere utilizzata per la riduzione dei dati in quanto consente di identificare un numero ridotto di valori che spiegano la maggior parte dei valori di varianza osservati in numerose variabili manifeste. L'analisi fattoriale può inoltre essere utilizzata per generare ipotesi relative a meccanismi causali oppure per esaminare le variabili per le analisi successive (ad esempio per identificare la collinearità prima di eseguire un'analisi di regressione lineare).

La procedura di analisi fattoriale permette un elevato grado di flessibilità:

- Sono disponibili sette metodi di estrazione fattoriale.
- Sono disponibili cinque metodi di rotazione, tra cui oblimin diretto e promax per le rotazioni non ortogonali.
- Sono disponibili tre metodi per il calcolo dei punteggi fattoriali, che possono essere salvati come variabili per le analisi successive.

Esempio. Quali sono gli atteggiamenti sottostanti che inducono le persone a rispondere a un'indagine politica in un determinato modo? Dall'esame delle correlazioni esistenti tra le voci dell'indagine risulta una significativa sovrapposizione tra diversi sottogruppi di voci. Ad esempio, le domande relative alle tasse tendono ad essere correlate fra loro, così come le domande relative alle questioni militari e così via. Grazie all'analisi fattoriale è possibile identificare il numero di fattori sottostanti e in molti casi determinare cosa rappresentano concettualmente tali fattori. È inoltre possibile calcolare i punteggi fattoriali per ciascun rispondente, un elemento che è possibile utilizzare in analisi successive. Ad esempio, è possibile creare un modello di regressione logistico che consenta di prevedere il comportamento di voto in base ai punteggi fattoriali.

Statistiche. Per ogni variabile: numero di casi validi, media e deviazione standard. Per ogni analisi del fattore: matrice di correlazione delle variabili, tra cui i livelli di significatività, determinante e inverso: matrice di correlazione riprodotta, tra cui anti-immagine; soluzione iniziale (comunalità, autovalori e percentuale di varianza spiegata); misura di Kaiser-Meyer-Olkin di adeguatezza del campionamento e test di sfericità di Bartlett; soluzione non ruotata, tra cui caricamenti dei fattori, comunalità e autovalori; soluzione ruotata, tra cui matrice del modello ruotato e matrice di trasformazione. Per le rotazioni oblique: modello ruotato e matrici della struttura; matrice del coefficiente di punteggio del fattore e matrice di covarianza del fattore. Grafici: grafico scree delle eguaglianze e grafico di caricamento dei primi due o tre fattori.

Considerazioni sui dati relative all'analisi fattoriale

Dati. Le variabili devono essere quantitative al livello di misura per *intervallo* o per *rapporto*. I dati categoriali (ad esempio religione o paese d'origine) non sono idonei per l'analisi fattoriale. I dati per cui è possibile calcolare i coefficienti di correlazione di Pearson sono idonei per l'analisi fattoriale.

Ipotesi. I dati devono avere una distribuzione normale bivariata per ogni coppia di variabili e le osservazioni devono essere indipendenti. Il modello di analisi fattoriale specifica che le variabili vengono determinate da fattori comuni (i fattori stimati dal modello) e fattori univoci (che non risultano sovrapposti tra le variabili osservate); le stime calcolate si basano sull'ipotesi che tutti i fattori univoci siano correlati reciprocamente e con i fattori comuni.

Per ottenere un'analisi fattoriale

1. Dai menu, scegliere:

   Analizza > **Riduzione delle dimensioni** > **Fattore...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare le variabili per l'analisi fattoriale.

Analisi fattoriale: Seleziona casi

Per selezionare i casi da usare nell'analisi:

1. Selezionare una variabile di selezione.
2. Fare clic su**Valore** per immettere un intero come variabile di selezione.

Nell'analisi fattoriale verranno utilizzati solo i casi con tale valore per la variabile di selezione.

Analisi fattoriale: Descrittive

Statistiche. Descrittive univariate include la media, la deviazione standard e il numero di casi validi per ogni variabile. Nella **soluzione iniziale** vengono visualizzate le comunalità iniziali, gli autovalori e la percentuale della varianza spiegata.

Matrice di correlazione. Le opzioni disponibili sono: coefficienti, livelli di significatività, determinante, inversa, riprodotta, anti-immagine, test KMO e test di sfericità di Bartlett.

- *Test KMO e test di sfericità di Bartlett*.La misura di adeguatezza del campionamento di Kaiser - Meyer - Olkin verifica se le correlazioni parziali tra le variabili sono piccole. Il test di sfericità di Bartlett verifica se la matrice di correlazione è una matrice identità, cosa che indicherebbe che il modello fattoriale è inappropriato.
- *Riprodotta*.La matrice di correlazione stimata dalla soluzione fattoriale. Vengono visualizzati anche i residui (differenze tra correlazioni stimate e osservate).
- *Anti-immagine*.La matrice di correlazione anti - immagine contiene i negativi dei coefficienti di correlazione parziale e la matrice di covarianza anti - immagine contiene i negativi delle covarianze parziali. In un buon modello fattoriale, la maggior parte degli elementi fuori dalla diagonale dovrebbe avere valori bassi. La misura dell'adeguatezza del campionamento per una variabile è visualizzata sulla diagonale della matrice di correlazione anti-immagine.

Analisi fattoriale: Estrazione

Metodo. Consente di specificare il metodo di estrazione del fattore. I metodi disponibili sono: componenti principali, minimi quadrati non pesati, minimi quadrati generalizzati, massima verosimiglianza, fattorizzazione dell'asse principale, fattorizzazione alfa e fattorizzazione immagine.

- *Analisi componenti principali*. Un metodo di estrazione dei fattori utilizzato per formare combinazioni lineari non correlate delle variabili osservate. La prima componente spiega la parte più alta di variabilità. Le componenti successive spiegano porzioni di variabilità decrescenti e sono tutte non correlate fra loro. L'analisi delle componenti principali viene usata per ottenere la soluzione fattoriale iniziale. Può essere usata quando una matrice di correlazione è singolare.
- *Metodo a minimi quadrati non ponderati*. Un metodo per l'estrazione dei fattori che minimizza la somma dei quadrati delle differenze tra la matrice di correlazione osservata e quella riprodotta, ignorando le diagonali.

- *Metodo a minimi quadrati generalizzati*. Un metodo di estrazione dei fattori che riduce al minimo la somma delle differenze al quadrato tra le matrici di correlazione osservate e riprodotte. Le correlazioni sono pesate tramite l'inverso della loro univocità, in modo da dare meno peso alle variabili con elevata univocità rispetto a quelle con univocità inferiore.
- *Metodo di massima verosimiglianza*. Un metodo di estrazione dei fattori che produce stime dei parametri che più probabilmente hanno prodotto la matrice di correlazione osservata se il campione proviene da una distribuzione normale multivariata. Le correlazioni sono pesate in base all'inverso dell'univocità delle variabili; viene usato un algoritmo iterativo.
- *Fattorizzazione dell'asse principale*. Un metodo di estrazione dei fattori dalla matrice di correlazione originale, con coefficienti di correlazione multipla al quadrato posti nella diagonale come stime iniziali delle comunalità. Questi caricamenti fattori vengono usati per stimare nuove comunalità che sostituiscono le vecchie stime sulla diagonale. Le iterazioni continuano fino a che le variazioni nelle comunalità da un'iterazione alla successiva soddisfano il criterio di convergenza per l'estrazione.
- *Fattorizzazione alfa*. Un metodo di estrazione dei fattori che considera le variabili nell'analisi come un campione dell'universo di variabili potenziali. Questo metodo massimizza l'affidabilità alfa dei fattori.
- *Fattorizzazione immagine*. Un metodo di estrazione dei fattori sviluppato da Guttman e basato sulla teoria delle immagini. La parte comune della variabile, detta immagine parziale, viene definita come la sua regressione lineare sulle variabili rimanenti, piuttosto che una funzione di fattori ipotetici.

Analizza. Consente di specificare una matrice di correlazione o una matrice di covarianza.

- **Matrice di correlazione.** Può risultare utile se le variabili dell'analisi vengono misurate su scale diverse.
- **Matrice di covarianza.** Può risultare utile se si intende applicare l'analisi fattoriale a più gruppi con varianze diverse per ogni variabile.

Estrai. È possibile mantenere tutti i fattori con autovalori superiori al valore specificato oppure mantenere solo il numero di fattori specificato.

Visualizza. Consente di richiedere la soluzione fattoriale non ruotata e un grafico scree.

- *Soluzione non ruotata*. Visualizza la matrice dei pesi fattoriali, le comunalità e gli autovalori della soluzione fattoriale.
- *Grafico scree*.Un grafico della varianza associata a ciascun fattore. Questo grafico viene usato per determinare il numero di fattori da mantenere. Di norma, il grafico mostra una distinta interruzione tra la brusca pendenza dei fattori di grandi dimensioni e la graduale sequenza del resto (lo scree).

Numero massimo di iterazioni per la convergenza. Consente di specificare il numero massimo di fasi che l'algoritmo può eseguire per valutare la soluzione.

Analisi fattoriale: Rotazione

Metodo. Consente di selezionare il metodo di rotazione dei fattori e se applicare la normalizzazione Kaiser. I metodi disponibili sono: varimax, equamax, quartimax, oblimin diretto e promax.

- *Varimax - metodo*. Un metodo di rotazione ortogonale che riduce al minimo il numero di variabili con elevati caricamenti su ciascun fattore. Questo metodo semplifica l'interpretazione dei fattori.
- *Metodo Oblimin diretto*.Un metodo per la rotazione obliqua (non ortogonale). Quando delta vale 0 (il valore predefinito), le soluzioni sono per la maggior parte oblique. Quando delta diventa negativo e aumenta in valore assoluto, i fattori cominciano a essere meno obliqui. Per sovrascrivere il delta predefinito di 0, immettere un numero inferiore o uguale a 0,8.
- *Metodo Quartimax*. Un metodo di rotazione che minimizza il numero di fattori necessari per spiegare ciascuna variabile. Questo metodo semplifica l'interpretazione delle variabili osservate.
- *Metodo di Equamax*. Un metodo di rotazione che è una combinazione del metodo varimax, che semplifica i fattori, e del metodo quartimax, che semplifica le variabili. Il numero di variabili con un notevole carico su un fattore e il numero di fattori necessario per spiegare una variabile sono minimizzati.

- *Rotazione Promax*. Una rotazione obliqua, che consente di correlare i fattori. Questa rotazione può essere calcolata più rapidamente di una rotazione Oblimin diretta ed è pertanto utile per dataset di grandi dimensioni.
- *Applica normalizzazione Kaiser*. Per impostazione predefinita, consente di applicare la normalizzazione Kaiser quando è specificata una rotazione.

Visualizza. Consente di includere l'output nella soluzione ruotata, nonché i grafici dei caricamenti per i primi due o tre fattori.

- *Soluzione ruotata*. Per ottenere una soluzione ruotata è necessario selezionare un metodo di rotazione. Per le rotazioni ortogonali vengono visualizzate la matrice ruotata dei modelli e la matrice di trasformazione. Per le rotazioni oblique vengono visualizzate la matrice dei modelli, la matrice di struttura e la matrice di correlazione dei fattori.
- *Grafico dei pesi fattoriali*. Grafico di caricamento tridimensionale dei primi tre fattori. Per le soluzioni a due fattori viene prodotto un grafico bidimensionale. Assente se l'analisi estrae un solo fattore. Se è stata richiesta la rotazione, il grafico visualizza la soluzione ruotata.

Numero massimo di iterazioni per la convergenza. Consente di specificare il massimo numero di fasi che l'algoritmo può eseguire per completare la rotazione.

Analisi fattoriale: Punteggi fattoriali

Salva come variabili. Consente di creare una nuova variabile per ciascun fattore nella soluzione finale.

Metodo. I metodi alternativi per il calcolo dei punteggi fattoriali sono regressione, Bartlett e Anderson-Rubin.

- *Metodo di regressione*. Un metodo per stimare i coefficienti del punteggio fattoriale. I punteggi prodotti hanno media 0 e varianza pari al quadrato della correlazione multipla fra i punteggi fattoriali stimati e i valori reali dei fattori. I punteggi possono essere correlati anche quando i fattori sono ortogonali.
- *Punteggi Bartlett*.Un metodo di stima dei coefficienti di punteggio fattoriale. I punteggi che vengono prodotti hanno una media di 0. La somma di quadrati dei fattori unici sulla gamma delle variabili è minimizzata.
- *Metodo Anderson-Rubin*. Un metodo per stimare i coefficienti del punteggio fattoriale; una modifica del metodo di Bartlett che garantisce l'ortogonalità dei fattori stimati. I punteggi prodotti hanno una media pari a 0, una deviazione standard pari a 1 e non sono correlati.

Visualizza matrice dei coefficienti di punteggio fattoriale. Mostra i coefficienti per cui vengono moltiplicate le variabili per ottenere i punteggi fattoriali. Vengono visualizzate anche le correlazioni tra i punteggi fattoriali.

Analisi fattoriale: Opzioni

Valori mancanti. Consente di specificare le modalità di gestione dei valori mancanti. Le scelte disponibili sono: Escludi casi *listwise*, Escludi casi *pairwise* o Sostituisci con la media.

Formato visualizzazione coefficienti. Consente di controllare alcuni aspetti delle matrici di output. È possibile ordinare i coefficienti per dimensioni ed eliminare i coefficienti con valori assoluti inferiori al valore specificato.

Funzioni aggiuntive del comando FACTOR

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare i criteri di convergenza per l'iterazione durante l'estrazione e la rotazione.
- Specificare i singoli grafici dei fattori ruotati.
- Specificare quanti punteggi fattoriali salvare.
- Specificare i valori diagonali per il metodo di calcolo dei fattori dell'asse principale.
- Scrivere le matrici di correlazione o le matrici dei fattori sul disco per poterle analizzare in seguito.

- Leggere e analizzare le matrici di correlazione o dei fattori.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Scelta di una procedura per il raggruppamento

È possibile eseguire cluster analysis utilizzando le procedure Analisi Cluster TwoStep, Cluster gerarchica o Cluster delle k-medie. Ogni procedura utilizza un algoritmo diverso per la creazione di cluster e include opzioni non disponibili nelle altre procedure.

Analisi cluster TwoStep. La procedura Analisi Cluster TwoStep risulta appropriata per molte applicazioni. Rende disponibili le seguenti funzioni univoche:

- Selezione automatica del numero di cluster più appropriato, oltre a misure per la scelta tra modelli di cluster.
- Possibilità di creare contemporaneamente due modelli di cluster basati su variabili categoriali e continue.
- Possibilità di salvare il modello di cluster in un file XML esterno e quindi di leggere tale file e aggiornare il modello di cluster utilizzando i dati più recenti.

Inoltre, questa procedura consente di analizzare file di dati di grandi dimensioni.

Analisi cluster gerarchica. La procedura Analisi cluster gerarchica è limitata a file di dati di dimensioni minori (ad esempio per il raggruppamento di centinaia di oggetti) e rende disponibili le seguenti funzioni univoche:

- Possibilità di raggruppare casi o variabili in cluster.
- Possibilità di calcolare un intervallo di soluzioni possibili e di salvare l'appartenenza a un cluster per ognuna di queste soluzioni.
- Diversi metodi per la formazione di cluster, la trasformazione delle variabili e la misurazione della dissimilarità tra cluster.

La procedura Analisi cluster gerarchica analizza variabili per intervallo (continue), binarie o di conteggio, purché le variabili siano tutte dello stesso tipo.

Analisi del cluster delle k-medie. La procedura Analisi del cluster delle k-medie è limitata a dati continui e richiede che il numero di cluster venga specificato anticipatamente, rendendo tuttavia disponibili le seguenti funzioni univoche:

- Possibilità di salvare le distanze dai centri di cluster per ogni oggetto.
- Possibilità di eseguire la lettura dai centri cluster iniziali e salvare i centri cluster finali in un file esterno IBM SPSS Statistica .

Inoltre, questa procedura consente di analizzare file di dati di grandi dimensioni.

Analisi cluster TwoStep

L'analisi cluster TwoStep è uno strumento di esplorazione che consente di rilevare raggruppamenti naturali, o cluster, all'interno di dataset, che non sarebbero altrimenti evidenti. L'algoritmo utilizzato da questa procedura presenta diverse funzioni che lo differenziano dalle tecniche di raggruppamento tradizionali:

- **Gestione di variabili categoriali e continue.** Se le variabili sono indipendenti, è possibile applicare una distribuzione normale multinomiale congiunta alle variabili categoriali e continue.
- **Selezione automatica del numero di cluster.** Mediante il confronto tra i valori dei criteri di scelta di modello appartenenti a diverse soluzioni di raggruppamento, la procedura è in grado di determinare automaticamente il numero ottimale di cluster.
- **Scalabilità.** Mediante la creazione di un albero delle funzioni cluster (o CF/cluster feature) che fornisce un riepilogo dei record, l'algoritmo TwoStep consente di analizzare file di dati di grandi dimensioni.

Esempio. Le aziende di prodotti retail e di consumo applicano regolarmente le tecniche di clustering ai dati che descrivono le abitudini di acquisto dei propri clienti, il sesso, l'età, il livello di reddito, ecc. Tali aziende adattano le loro strategie di sviluppo dei prodotti e di marketing a ciascun gruppo di consumo per incrementare le vendite e creare una fidelizzazione al marchio.

Misura di distanza. Questa selezione determina la modalità di calcolo della similarità tra due cluster.

- **Logaritmo della verosimiglianza.** La misura di verosimiglianza applica una distribuzione per probabilità alle variabili. Si suppone che le variabili continue vengano distribuite normalmente, mentre le variabili categoriali in base al modello multinomiale. Si suppone che tutte le variabili siano indipendenti.
- **Euclidea.** La misura euclidea è la distanza in "linea retta" tra due cluster. Può essere utilizzata solo quando tutte le variabili sono continue.

Numero di cluster. Questa selezione consente di specificare la modalità di definizione del numero dei cluster.

- **Determina automaticamente.** La procedura determina automaticamente il numero di cluster ottimale, mediante i criteri specificati nel gruppo Criteri di raggruppamento. È inoltre possibile immettere un intero positivo per definire il numero di cluster massimo che la procedura dovrà prendere in considerazione.
- **Specifica fisso.** Consente di definire un numero fisso di cluster nella soluzione. Immettere un intero positivo.

Conteggio di variabili continue. Questo gruppo fornisce un riepilogo delle opzioni di standardizzazione relative alle variabili continue specificate nella finestra di dialogo Opzioni. Per ulteriori informazioni, consultare l'argomento “Opzioni di Analisi cluster TwoStep” a pagina 222.

Criterio di clustering. Questa selezione determina la modalità di definizione del numero dei cluster mediante l'algoritmo di cluster automatico. È possibile specificare il modello Criterio bayesiano di Schwarz (BIC, Bayesian Information Criterion) o il modello Criterio di informazione di Akaike (AIC, Akaike Information Criterion).

Considerazioni sui dati dell'analisi cluster TwoStep

Dati. Questa procedura può essere utilizzata sia con le variabili continue sia con le variabili categoriali. I casi rappresentano gli oggetti da raggruppare e le variabili corrispondono agli attributi in base ai quali viene eseguito il raggruppamento.

Ordine dei casi. L'albero delle funzioni cluster e la soluzione finale possono dipendere dall'ordine dei casi. Per ridurre al minimo gli effetti dell'ordine, disporre i casi in ordine casuale. Può essere utile ottenere più soluzioni diverse con casi disposti in ordini casuali diversi per verificare la stabilità di una soluzione specifica. Nei casi in cui questa operazione è complessa a causa delle dimensioni eccessive dei file, è possibile effettuare più operazioni con un campione di casi disposti in ordini casuali diversi.

Ipotesi. La misura della distanza della verosimiglianza assume che le variabili nel modello di cluster siano indipendenti. A ogni variabile continua inoltre si suppone inoltre che sia associata una distribuzione normale o gaussiana, mentre a ogni variabile categoriale una distribuzione multinomiale. La verifica empirica interna indica che la procedura è piuttosto robusta rispetto alle violazioni sia delle ipotesi di indipendenza sia delle ipotesi di distribuzione, ma è consigliabile verificare fino a che punto tali ipotesi vengano soddisfatte.

Utilizzare la procedura Correlazioni bivariate per testare l'indipendenza di due variabili continue. Utilizzare la procedura Tabelle di contingenza per verificare l'indipendenza di due variabili categoriali. Utilizzare la procedura Medie per testare l'indipendenza tra una variabile continua e la variabile categoriale. Utilizzare la procedura Esplora per testare la normalità di una variabile continua. Utilizzare la procedura Test del chi-quadrato per testare se una variabile categoriale ha una distribuzione multinomiale specificata.

Per ottenere un'analisi cluster TwoStep

1. Dai menu, scegliere:

   Analizza > **Classifica** > **TwoStep Cluster...**

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una o più variabili categoriali o continue.

Facoltativamente, è possibile:

- Adattare i criteri in base ai quali sono stati creati i cluster.
- Selezionare le impostazioni di gestione del rumore, allocazione di memoria, standardizzazione delle variabili e input del modello di cluster.
- Richiedere l'output del Visualizzatore modelli.
- Salvare i risultati del modello nel file di lavoro o in un file XML esterno.

Opzioni di Analisi cluster TwoStep

Trattamento valori anomali. Questo gruppo consente di considerare i valori anomali in particolare durante il clustering, se l'albero delle funzioni cluster (CF) si riempie. L'albero delle funzioni cluster (CF) è pieno quando non è più in grado di accettare casi in un nodo foglia e nessun nodo foglia può essere suddiviso.

- Se si seleziona la gestione del rumore e l'albero delle funzioni cluster (CF) risulta pieno, sarà possibile ampliarlo posizionando i casi mal distribuiti in più foglie all'interno di una foglia specifica per il "rumore". Una foglia contiene casi mal distribuiti quando il numero dei casi è inferiore alla percentuale specificata per la dimensione massima della foglia. Dopo aver ampliato l'albero, i valori anomali vengono inseriti nell'albero delle funzioni cluster (CF), se possibile. Altrimenti, vengono eliminati.
- Se non si seleziona la gestione del rumore e l'albero delle funzioni cluster (CF) risulta pieno, sarà possibile ampliarlo utilizzando una soglia di modifica della distanza più elevata. Dopo il raggruppamento finale, i valori non assegnati a un cluster vengono definiti valori anomali. Al cluster di valori anomali viene assegnato un numero di identificazione –1 e non viene incluso nel conteggio del numero di cluster.

Allocazione di memoria. Questo gruppo consente di specificare in megabyte (MB) la quantità massima di memoria che l'algoritmo del cluster può utilizzare. Se questa quantità massima viene superata, la procedura utilizzerà il disco per memorizzare le informazioni che non è possibile inserire nella memoria. Specificare un numero maggiore o uguale a 4.

- Per informazioni sul valore massimo per il sistema, rivolgersi all'amministratore di sistema.
- L'algoritmo potrebbe non riuscire a trovare il numero di cluster corretto o specificato se questo valore è troppo basso.

Standardizzazione della variabile. L'algoritmo di cluster funziona con variabili continue standardizzate. Qualsiasi variabile non standardizzata deve essere impostata come "Da standardizzare". Per risparmiare tempo e calcoli, è possibile impostare le variabili continue già standardizzate come "Già standardizzate".

Opzioni avanzate

Criteri di ottimizzazione dell'albero delle funzioni cluster (CF). Le seguenti impostazioni dell'algoritmo di cluster riguardano in modo specifico l'albero delle funzioni cluster (CF) e devono essere modificate con cautela:

- **Soglia modifica distanza iniziale.** Si tratta della soglia iniziale utilizzata per ampliare l'albero delle funzioni cluster (CF). Se dopo l'inserimento di un determinato caso in una foglia dell'albero delle funzioni cluster (CF), la distanza risulta inferiore alla soglia, la foglia non viene suddivisa. Se la distanza supera la soglia, la foglia può essere suddivisa.
- **Ramificazioni massime (per nodo foglia).** Il numero massimo di nodi figlio per un nodo foglia.
- **Massima profondità dell'albero.** Il numero massimo di livelli dell'albero delle funzioni cluster (CF).
- **Massimo numero di nodi possibile.** Indica il numero massimo di nodi dell'albero delle funzioni cluster (CF) che questa procedura è in grado di generare, in base alla funzione $(b^{d+1} – 1) / (b – 1)$, dove b rappresenta le ramificazioni massime e d è la profondità massima dell'albero. Tenere presente che un albero delle funzioni cluster (CF), di dimensioni eccessive può rappresentare un peso considerevole per

le risorse del sistema e compromettere quindi la prestazione della procedura. Ogni nodo richiede un minimo di 16 byte.

Aggiornamento del modello di cluster. Questo gruppo consente di importare e aggiornare un modello di cluster generato da un'analisi precedente. Il file di input contiene l'albero delle funzioni cluster (CF) in formato XML. Il modello viene quindi aggiornato con i dati nel file attivo. È necessario selezionare i nomi delle variabili nella finestra di dialogo principale in base allo stesso ordine dell'analisi precedente. Il file XML non viene modificato, a meno che le nuove informazioni relative al modello non vengano inserite nello stesso file. Per ulteriori informazioni, consultare l'argomento "Output di Analisi cluster TwoStep" a pagina 223.

Se viene selezionato l'aggiornamento del modello di cluster, verranno utilizzate le opzioni per la generazione dell'albero delle funzioni cluster (CF) specificate per il modello originale. Vengono quindi utilizzate le impostazioni del modello salvato relative a misura della distanza, gestione del rumore, allocazione di memoria e ottimizzazione dell'albero delle funzioni cluster (CF), mentre qualsiasi nuova impostazione specificata nelle finestre di dialogo viene ignorata.

Nota: quando si esegue un aggiornamento del modello di cluster, la procedura presume che nessuno dei casi selezionati nel dataset attivo sia stato utilizzato per creare il modello di cluster originale. La procedura si basa inoltre sul presupposto che i casi utilizzati nell'aggiornamento del modello di cluster provengono dalla stessa popolazione di casi utilizzati per creare il modello originale, le medie e le varianze delle variabili continue e i livelli delle variabili categoriali devono quindi essere le stesse per i due insiemi di casi. Se gli insiemi di casi precedenti e correnti provengono da una popolazione eterogenea, è necessario eseguire la procedura Analisi cluster TwoStep negli insiemi di casi combinati per ottenere risultati ottimali.

Output di Analisi cluster TwoStep

Output. Questo gruppo fornisce le opzioni per la visualizzazione dei risultati dei raggruppamenti.

- **Tabelle pivot.** I risultati vengono visualizzati nelle tabelle pivot.
- **Grafici e tabelle in Visualizzatore modelli.** I risultati vengono visualizzati nel Visualizzatore modelli.
- **Campi valutazione.** Calcola i dati dei cluster per le variabili che non sono stati utilizzati nella creazione dei cluster. I campo valutazione possono essere visualizzati insieme alle funzioni di input nel Visualizzatore modelli selezionandoli nella finestra di dialogo secondaria Visualizza. I campi con valori mancanti vengono ignorati.

File di dati di lavoro. Questo gruppo consente di salvare le variabili all'interno del dataset attivo.

- **Crea variabile di appartenenza cluster.** Questa variabile contiene un numero di identificazione del cluster per ogni caso. Il nome di questa variabile è *tsc_n*, dove *n* è un intero positivo che indica l'ordinale dell'operazione di salvataggio del dataset attivo eseguita mediante questa procedura in una determinata sessione.

File XML. Il modello di cluster finale e l'albero delle funzioni cluster (CF) rappresentano due tipi di file di output che possono essere esportati in formato XML.

- **Esporta modello finale.** Il modello di cluster finale viene esportato nel file specificato in formato XML (PMML). È possibile utilizzare questo file modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio.
- **Esporta albero delle funzioni cluster (CF).** Questa opzione consente di salvare lo stato attuale dell'albero del cluster e di aggiornarlo successivamente utilizzando i dati più recenti.

Il Visualizzatore cluster

I modelli di cluster in genere vengono utilizzati per trovare gruppi (o cluster) di record simili in base alle variabili esaminate, dove la somiglianza tra i membri dello stesso gruppo è elevata e la somiglianza tra i membri di gruppi differenti è bassa. I risultati possono essere utilizzati per identificare le associazioni che altrimenti non sarebbero visibili. Ad esempio, mediante l'analisi cluster delle preferenze dei clienti, del livello di reddito e delle abitudini di acquisto è possibile identificare i tipi di clienti che risponderanno con maggiore probabilità a una determinata campagna di marketing.

Sono disponibili due approcci per interpretare i risultati in una visualizzazione cluster.

- Esaminare i cluster per determinare le caratteristiche univoche del cluster in questione. *Un cluster contiene tutti coloro che chiedono denaro in prestito con reddito elevato? Questo cluster contiene più record di altri?*
- Esaminare i campi nei cluster per determinare come sono distribuiti i valori nei cluster. *Il livello di istruzione determina l'appartenenza a un cluster? Un punteggio di credito elevato distingue tra l'appartenenza a un cluster o a un altro?*

Utilizzando le viste principali e le varie viste collegate nel Visualizzatore cluster, è possibile ottenere informazioni utili per rispondere a queste domande.

Per informazioni sul modello di cluster, attivare (fare doppio clic) l'oggetto Visualizzatore modelli nel Visualizzatore.

Visualizzatore cluster

Il Visualizzatore cluster è composto da due pannelli, la vista principale sul lato sinistro e la vista collegata, o ausiliaria, sul lato destro. Vi sono due viste principali:

- Riepilogo del modello (valore predefinito). Per ulteriori informazioni, consultare l'argomento “Vista Riepilogo del modello” a pagina 224.
- Raggruppamenti. Per ulteriori informazioni, consultare l'argomento “Vista Cluster” a pagina 225.

Le viste collegate/ausiliarie sono quattro:

- Importanza predittore. Per ulteriori informazioni, consultare l'argomento “Visualizzazione Importanza predittore nei cluster” a pagina 226.
- Dimensioni cluster (valore predefinito). Per ulteriori informazioni, consultare l'argomento “Vista Dimensioni cluster” a pagina 226.
- Distribuzione cella. Per ulteriori informazioni, consultare l'argomento “Vista Distribuzione delle celle” a pagina 227.
- Confronto tra cluster. Per ulteriori informazioni, consultare l'argomento “Vista Confronto tra cluster” a pagina 227.

Vista Riepilogo del modello

La vista Riepilogo del modello mostra un'istantanea, o riepilogo, del modello di cluster, inclusa una misura Silhouette della coesione e separazione del cluster, che è ombreggiata per indicare risultati insufficienti, sufficienti o buoni. Questa istantanea consente di controllare rapidamente se la qualità è insufficiente, nel qual caso è possibile decidere di ritornare al nodo di modellazione per modificare le impostazioni del modello di cluster per ottenere un risultato migliore.

I risultati insufficienti, sufficienti e buoni si basano sul lavoro di Kaufman e Rousseeuw (1990) in merito all'interpretazione delle strutture di cluster. Nella vista Riepilogo del modello un buon risultato equivale ai dati che rappresentano la valutazione di Kaufman e Rousseeuw come prova ragionevole o forte della struttura del cluster, un risultato sufficiente rappresenta la loro valutazione della prova debole e un risultato insufficiente rappresenta la loro valutazione della prova non significativa.

La misura Silhouette calcola la media in tutti i record, (B–A) / max(A,B), in cui A è la distanza del record dal centro del suo cluster e B è la distanza del record dal centro del cluster più vicino a cui non appartiene. Il coefficiente silhouette pari a 1 indica che tutti i casi si trovano direttamente nei centri dei loro cluster. Il valore –1 indica che tutti i casi si trovano nei centri di altri cluster. Il valore 0 indica, in media, che i casi sono equidistanti tra il centro del proprio cluster e l'altro cluster più vicino.

Il riepilogo include una tabella contenente le informazioni riportate di seguito.

- **Algoritmo.** L'algoritmo di clustering utilizzato, ad esempio, "TwoStep".
- **Funzioni di input.** Il numero di campi, noti anche come **input** o **predittori**.
- **Cluster.** Il numero di cluster nella soluzione.

Vista Cluster

La vista Cluster contiene una griglia di cluster per funzioni che include i nomi, le dimensioni e i profili di ciascun cluster.

Le colonne della griglia contengono le informazioni riportate di seguito.

- **Cluster.** I numeri di cluster creati dall'algoritmo.
- **Etichetta.** Le etichette applicate a ciascun cluster (per impostazione predefinita, l'etichetta è vuota). Fare doppio clic nella cella per immettere un'etichetta che descriva il contenuto del cluster; ad esempio, "Acquirenti di automobili di lusso".
- **Descrizione.** Una descrizione del contenuto del cluster (per impostazione predefinita, la descrizione è vuota). Fare doppio clic nella cella per immettere una descrizione del cluster; ad esempio "Professionisti cinquantacinquenni con un reddito superiore a $100.000".
- **Dimensione.** La dimensione di ogni cluster come percentuale del campione di cluster globale. Ogni cella di dimensione all'interno della griglia visualizza una barra verticale che mostra la percentuale della dimensione all'interno del cluster, una percentuale della dimensione in formato numerico e i conteggi dei casi del cluster.
- **Funzioni.** I predittori e gli input singoli, ordinati, per impostazione predefinita, in base all'importanza globale. Se delle colonne hanno la stessa dimensione, vengono visualizzate in ordine crescente dei numeri di cluster.

  L'importanza globale delle funzioni viene indicata dal colore dell'ombreggiatura dello sfondo della cella; la funzione più importante è la più scura; la funzione meno importante non è ombreggiata. Una guida sopra la tabella indica l'importanza collegata a ogni colore della cella della funzione.

Quando si passa il mouse su una cella, vengono visualizzati il nome completo/l'etichetta della funzione e il valore relativo all'importanza della cella. È possibile che vengano visualizzate altre informazioni, a seconda del tipo di vista e di funzione. Nella vista Centri cluster sono incluse le statistiche della cella e il valore della cella; ad esempio: “Media: 4.32”. Per le funzioni categoriali la cella mostra il nome della categoria più frequente (modale) e la relativa percentuale.

Nella vista Cluster è possibile selezionare diversi modi per visualizzare le informazioni del cluster:

- Trasporre cluster e funzioni. Per ulteriori informazioni, consultare l'argomento “Trasponi cluster e funzioni” a pagina 225.
- Ordinare le funzioni. Per ulteriori informazioni, consultare l'argomento “Ordina funzioni” a pagina 225.
- Ordinare i cluster. Per ulteriori informazioni, consultare l'argomento “Ordina cluster” a pagina 226.
- Selezionare il contenuto della cella. Per ulteriori informazioni, consultare l'argomento “Contenuto della cella” a pagina 226.

Trasponi cluster e funzioni

Per impostazione predefinita, i cluster vengono visualizzati come colonne e le funzioni vengono visualizzate come righe. Per invertire questa visualizzazione, fare clic sul pulsante **Trasponi cluster e funzioni** sul lato destro dei pulsanti **Ordina funzioni per**. Ad esempio, questa operazione è utile se sono visualizzati molti cluster e si desidera ridurre lo scorrimento orizzontale necessario per visualizzare i dati.

Ordina funzioni

I pulsanti **Ordina funzioni per** consentono di selezionare come vengono visualizzate le celle delle funzioni.

- **Importanza globale.** Si tratta dell'ordinamento predefinito. Le funzioni vengono disposte in ordine decrescente di importanza globale e l'ordinamento è lo stesso nei cluster. Se delle funzioni hanno valori di importanza correlati, le funzioni correlate vengono elencate in ordine crescente dei nomi delle funzioni.
- **Importanza entro i cluster.** Le funzioni vengono ordinate rispetto alla loro importanza per ogni cluster. Se delle funzioni hanno valori di importanza correlati, le funzioni correlate vengono elencate in ordine

crescente dei nomi delle funzioni. Quando si sceglie questa opzione, l'ordinamento in genere varia tra i cluster.

- **Nome.** Le funzioni vengono disposte per nome in ordine alfabetico.
- **Ordine dei dati.** Le funzioni vengono disposte in base al loro ordine nel dataset.

Ordina cluster

Per impostazione predefinita, i cluster sono disposti in ordine decrescente di dimensione. I pulsanti **Ordina cluster per** consentono di disporli per nome in ordine alfabetico o, se sono state create delle etichette univoche, in ordine alfanumerico di etichetta.

Le funzioni che hanno la stessa etichetta sono ordinate per nome cluster. Se i cluster sono ordinati per etichetta e si modifica l'etichetta di un cluster, l'ordinamento viene aggiornato automaticamente.

Contenuto della cella

I pulsanti **Celle** consentono di modificare la visualizzazione del contenuto della cella per i campi delle funzioni e di valutazione.

- **Centri cluster.** Per impostazione predefinita, le celle visualizzano i nomi/le etichette delle funzioni e la tendenza centrale per ogni combinazione di cluster/funzione. Per i campi continui viene visualizzata la media e per i campi categoriali viene mostrata la moda (la categoria riscontrata più di frequente) con la percentuale della categoria.
- **Distribuzioni assolute.** Mostra le distribuzioni assolute dei nomi/delle etichette delle funzioni all'interno di ogni cluster. Per le funzioni categoriali, i grafici a barre vengono mostrati sovrapposti e le categorie sono disposte in ordine crescente dei valori dei dati. Per le funzioni continue, viene mostrato un grafico di densità livellato, che utilizza gli stessi endpoint e intervalli per ogni cluster.

  Il pannello rosso scuro mostra la distribuzione del cluster, mentre il pannello più chiaro rappresenta i dati globali.

- **Distribuzioni relative.** Mostra i nomi/le etichette della funzione e le distribuzioni relative nelle celle. In generale, le visualizzazioni sono simili a quelle mostrate per le distribuzioni assolute, con la differenza che vengono visualizzate le distribuzioni relative.

  Il pannello rosso scuro mostra la distribuzione del cluster, mentre il pannello più chiaro rappresenta i dati globali.

- **Vista di base.** Quando sono presenti molti cluster, può essere difficile vedere tutti i dettagli senza scorrere i dati. Per ridurre la necessità dello scorrimento, selezionare questa vista per modificare la visualizzazione su una versione più compatta della tabella.

Visualizzazione Importanza predittore nei cluster

La visualizzazione Importanza predittore mostra l'importanza relativa di ciascun campo nella stima del modello.

Vista Dimensioni cluster

La vista Dimensioni cluster mostra un grafico a torta contenente ogni cluster. La dimensione della percentuale di ogni cluster viene visualizzata in ogni sezione; passare il mouse su ogni sezione per visualizzare il conteggio nella sezione in questione.

Una tabella sotto il grafico elenca le informazioni sulla dimensione riportate di seguito.

- La dimensione del cluster più piccolo (il conteggio e la percentuale dell'insieme).
- La dimensione del cluster più grande (il conteggio e la percentuale dell'insieme).
- Il rapporto della dimensione del cluster più grande con il cluster più piccolo.

Vista Distribuzione delle celle

La vista Distribuzione delle celle mostra un grafico espanso e più dettagliato della distribuzione dei dati per le celle della funzione selezionate nella tabella nel pannello principale Cluster.

Vista Confronto tra cluster

La vista Confronto tra cluster è composta da un layout stile griglia con le funzioni nelle righe e i cluster selezionati nelle colonne. Questa vista consente di comprendere meglio i fattori che costituiscono i cluster; inoltre, consente di vedere le differenze tra i cluster non solo rispetto ai dati globali, ma tra di loro.

Per selezionare i cluster da visualizzare, fare clic nella parte superiore della colonna del cluster nel pannello principale Cluster. Utilizzare i tasti Ctrl-clic o Maiusc-clic per selezionare o deselezionare più di un cluster per il confronto.

Nota: è possibile selezionare fino a cinque cluster per la visualizzazione.

I cluster vengono mostrati nell'ordine in cui sono stati selezionati, mentre l'ordine dei campi è determinato dall'opzione **Ordina funzioni per**. Quando si seleziona **Importanza entro i cluster**, i campi vengono sempre ordinati per importanza globale.

I grafici di sfondo mostrano le distribuzioni globali di ciascuna funzione.

- Le funzioni categoriali vengono mostrate come grafici a punti, in cui la dimensione del punto indica la categoria più frequente/modale per ciascun cluster (per funzione).
- Le funzioni continue vengono visualizzate come grafici a scatole, che mostrano le mediane globali e gli intervalli interquartili.

Sovrapposti a queste viste di sfondo sono i grafici a scatole per i cluster selezionati:

- Per le funzioni continue, i contrassegni a punta quadrata e le linee orizzontali indicano la mediana e l'intervallo interquartile per ciascun cluster.
- Ogni cluster è rappresentato da un colore differente, mostrato nella parte superiore della vista.

Navigazione nel visualizzatore cluster

Il visualizzatore cluster è una visualizzazione interattiva. È possibile:

- Selezionare un campo o un cluster per visualizzare più dettagli.
- Confrontare i cluster per selezionare gli elementi a cui si è interessati.
- Modificare la visualizzazione.
- Trasporre gli assi.

Utilizzo delle barre degli strumenti

L'utente controlla le informazioni visualizzate nei pannelli sulla destra e sulla sinistra utilizzando le opzioni della barra degli strumenti. È possibile modificare l'orientamento della visualizzazione (dall'alto verso il basso, da sinistra a destra o da destra a sinistra) utilizzando i controlli della barra degli strumenti. Inoltre, è possibile ripristinare le impostazioni predefinite del visualizzatore e aprire una finestra di dialogo per specificare il contenuto della vista Cluster nel pannello principale.

Le opzioni **Ordina funzioni per**, **Ordina cluster per**, **Celle** e **Visualizzazione** sono disponibili solo quando si seleziona la vista **Cluster** nel pannello principale. Per ulteriori informazioni, consultare l'argomento “Vista Cluster” a pagina 225.

Tabella 4. Icone della barra degli strumenti

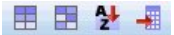
Icona	Argomento
	Consultare Trasponi cluster e funzioni
	Consultare Ordina funzioni per

Tabella 4. Icone della barra degli strumenti (Continua)	
Icona	**Argomento**
	Consultare Ordina cluster per
	Consultare Celle

Controllo della visualizzazione della vista cluster

Per controllare cosa viene visualizzato nella vista Cluster nel pannello principale, fare clic sul pulsante **Visualizza**; viene aperta la finestra Visualizza.

Funzioni. Opzione selezionata per impostazione predefinita. Per nascondere tutte le funzioni di input, deselezionare la casella di controllo.

Campi valutazione. Scegliere i campi di valutazione (campi non utilizzati per creare il modello cluster, ma inviati al visualizzatore modelli per valutare i cluster) da visualizzare; nessuno di essi viene visualizzato per impostazione predefinita. *Nota:* il campo di valutazione deve essere una stringa con più di un valore. Questa casella di controllo non è disponibile se non è disponibile alcun campo di valutazione.

Descrizioni cluster. Opzione selezionata per impostazione predefinita. Per nascondere tutte le celle di descrizione, deselezionare la casella di controllo.

Dimensioni cluster. Opzione selezionata per impostazione predefinita. Per nascondere tutte le celle di dimensione del cluster, deselezionare la casella di controllo.

Numero massimo di categorie. Specificare il numero massimo di categorie da visualizzare nei grafici delle funzioni categoriali; il valore predefinito è 20.

Filtraggio dei record

Se si desidera avere ulteriori informazioni sui casi in un determinato cluster o gruppo di cluster, è possibile selezionare un sottoinsieme di record per un'analisi più approfondita basata sui cluster selezionati.

1. Selezionare i cluster nella vista Cluster del Visualizzatore cluster. Per selezionare più cluster, utilizzare Ctrl-clic.
2. Dai menu, scegliere:

   Generate > **Record di filtro ...**
3. Inserire un nome per la variabile di filtro. I record dai cluster selezionati riceveranno un valore 1 per questo campo. Tutti gli altri record riceveranno il valore 0 e verranno esclusi dalle analisi successive finché non si modifica lo stato del filtro.
4. Fare clic su **OK**.

Analisi cluster gerarchica

Questa procedura consente di identificare gruppi di casi relativamente omogenei in base alle caratteristiche selezionate, utilizzando un algoritmo che inizia con ciascun caso (o variabile) in un cluster distinto e che combina i cluster fino a quando ne rimane solo uno. È possibile analizzare le variabili grezze oppure scegliere una delle trasformazioni di standardizzazione disponibili. Le misure di similarità e dissimilarità vengono generate dalla procedura Prossimità. A ciascun livello verranno visualizzate statistiche in base alle quali selezionare la soluzione migliore.

Esempio. Esistono gruppi di trasmissioni televisive identificabili che attraggono tipi di audience analoghi all'interno di ciascun gruppo? Utilizzando l'analisi cluster gerarchica è possibile raggruppare le trasmissioni televisive (casi) in gruppi omogenei in base alle caratteristiche degli spettatori. Questo metodo può essere utilizzato per identificare i segmenti di mercato. In alternativa, è possibile raggruppare le città (casi) in gruppi omogenei in modo che da poter selezionare città con caratteristiche confrontabili per verificare diverse strategie di mercato.

Statistiche. Pianificazione di agglomerazione, matrice delle distanze (o similarità) e cluster di appartenenza per un'unica soluzione o una serie di soluzioni. Grafici: Dendrogrammi e grafici a ghiacciolo.

Considerazioni sui dati per l'Analisi cluster gerarchica

Dati. Le variabili possono essere quantitative, binarie o dati di conteggio. Lo scaling delle variabili è molto importante in quanto le differenze di scaling possono influire sulle soluzioni cluster. Se lo scaling delle variabili presenta differenze notevoli (ad esempio, una variabile viene misurata in dollari e l'altra in anni), è consigliabile standardizzarle. Ciò può essere effettuato in modo automatico mediante la procedura Analisi cluster gerarchica.

Ordine dei casi. Se le distanze assegnate o le similarità sono presenti nei dati di input o nei cluster aggiornati durante l'unione, la soluzione cluster risultante può essere influenzata dall'ordine dei casi del file. Può essere utile ottenere più soluzioni diverse con casi disposti in ordini casuali diversi per verificare la stabilità di una soluzione specifica.

Ipotesi. Le misure di dissimilarità o di similarità utilizzate devono essere idonee per i dati analizzati. Per ulteriori informazioni sulla scelta delle misure di dissimilarità e similarità, vedere la procedura Prossimità. È inoltre necessario includere nell'analisi tutte le variabili significative. L'omissione di variabili importanti può portare a soluzioni improprie. Poiché l'analisi cluster gerarchica rappresenta un metodo esplorativo, i risultati devono essere considerati provvisori finché non vengano confermati da un campione indipendente.

Per ottenere una Analisi cluster gerarchica

1. Dai menu, scegliere:

   Analizza > **Classifica** > **Cluster gerarchico...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Per raggruppare i casi in cluster è necessario selezionare almeno una variabile numerica. Per raggruppare le variabili in cluster è necessario selezionare almeno tre variabili numeriche.

È inoltre possibile selezionare una variabile di identificazione per etichettare i casi.

Analisi cluster gerarchica: Metodo

Metodo cluster. Le alternative disponibili sono: Legame medio fra i gruppi, Legame medio entro gruppi, Del vicino più vicino, Del vicino più lontano, Baricentro, Mediana e Ward.

Misura. Consente di specificare la misura di similarità o dissimilarità da utilizzare per il raggruppamento. Selezionare il tipo di dati e la misura di similarità o dissimilarità desiderata:

- **Intervallo.** Le alternative disponibili sono: Distanza euclidea, Distanza euclidea al quadrato, Coseno, Correlazione di Pearson, Chebychev, City-Block, Minkowski e Personalizzato.
- **Conteggi.** Le alternative disponibili sono: Misura chi-quadrato e Misura phi-quadrato.
- **Binaria.** Le alternative disponibili sono: Distanza euclidea, Distanza euclidea al quadrato, Differenza di dimensione, Differenza di modello, Varianza, Dispersione, Forma, Corrispondenza semplice, Correlazione phi a 4 punti, Lambda, *D* di Anderberg, Dice, Hamann, Jaccard, Kulczynski 1, Kulczynski 2, Lance e Williams, Ochiai, Rogers e Tanimoto, Russel e Rao, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Sokal e Sneath 4, Sokal e Sneath 5, *Y* di Yule e *Q*di Yule.

Trasforma valori. Consente di standardizzare i valori dei dati per casi o valori prima di calcolare le prossimità (non disponibile per i dati binari). I metodi di standardizzazione disponibili sono Punteggi *z*, Intervallo da –1 a 1, Intervallo da 0 a 1, Grandezza massima di 1, Media di 1 e Deviazione standard di 1.

Trasforma misure. Consente di trasformare i valori generati dalla misura della distanza. Questi verranno applicati dopo il calcolo della misura della distanza. Le alternative disponibili sono Valori assoluti, Cambia segno e Modifica scala su intervallo 0-1.

Analisi cluster gerarchica: Statistiche

Pianificazione di agglomerazione. Consente di visualizzare i casi o i cluster accorpati ad ogni stadio, le distanze tra i casi o i cluster da accorpare e l'ultimo livello di cluster in cui un caso (o una variabile) è stato accorpato al cluster.

Matrice di prossimità. Fornisce le distanze o le similarità tra gli elementi.

Appartenenza cluster. Viene visualizzato il cluster a cui viene assegnato ciascun caso a uno o più stadi della combinazione dei cluster. Le opzioni disponibili sono Soluzione unica e Intervallo di soluzioni.

Analisi cluster gerarchica: Grafici

Dendrogramma. Viene visualizzato un *dendrogramma*. Utilizzando i dendrogrammi è possibile valutare la coesione dei cluster formati ed ottenere informazioni sul numero di cluster che è opportuno tenere.

Stalattite. Visualizza un *grafico a ghiacciolo*, inclusi tutti i cluster o l'intervallo di cluster specificato. Nei grafici a ghiacciolo vengono visualizzate informazioni sulle modalità con cui i casi vengono combinati in cluster ad ogni iterazione dell'analisi. Specificando l'orientamento desiderato è possibile selezionare un grafico verticale o orizzontale.

Analisi cluster gerarchica: Salva nuove variabili

Appartenenza cluster. Consente di salvare i cluster di appartenenza per una soluzione unica o per un intervallo di soluzioni. Le variabili salvate possono essere utilizzate in analisi successive per valutare altre differenze tra i gruppi.

Funzioni aggiuntive della sintassi del comando CLUSTER

La procedura Cluster gerarchica usa la sintassi del comando CLUSTER. Il linguaggio della sintassi dei comandi consente inoltre di:

- Usare più metodi di raggruppamento in una singola analisi.
- Leggere ed analizzare una matrice di prossimità.
- Scrivere una matrice di prossimità sul disco per analizzarla in seguito.
- Specificare i valori per la potenza e la radice nella misura della distanza personalizzata (potenza).
- Specificare i nomi delle variabili salvate.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Analisi del cluster delle k-medie

Questa procedura consente di identificare gruppi di casi relativamente omogenei in base alle caratteristiche selezionate, utilizzando un algoritmo in grado di gestire un elevato numero di casi. Tale algoritmo, tuttavia, richiede l'indicazione del numero di cluster. È possibile specificare i centri iniziali del cluster, se si conoscono queste informazioni. È possibile selezionare uno dei due metodi disponibili per la classificazione dei casi, ovvero l'aggiornamento iterativo dei centri cluster oppure la semplice classificazione. È possibile salvare l'appartenenza al cluster, le informazioni sulla distanza e i centri del cluster finali. È inoltre possibile specificare una variabile i cui valori possono essere utilizzati per etichettare l'output casewise. i può inoltre richiedere l'analisi delle statistiche *F* di varianza. Se da un lato queste statistiche sono opportunistiche, ovvero vengono eseguiti tentativi di raggruppamenti che presentino differenze, le corrispondenti dimensioni relative forniscono informazioni sul contributo apportato da ciascuna variabile alla separazione dei gruppi.

Esempio. Quali sono i gruppi di show televisivi che attraggono un pubblico analogo all'interno di ciascun gruppo? Il metodo cluster *k*-medie consente di raggruppare gli show televisivi (casi) in*k* gruppi omogenei in base alle caratteristiche degli spettatori. Questo processo può essere utilizzato per identificare i segmenti di mercato. In alternativa, è possibile raggruppare le città (casi) in gruppi omogenei in modo che da poter selezionare città con caratteristiche confrontabili per verificare diverse strategie di mercato.

Statistiche. Soluzione completa: centri cluster iniziali, tabella ANOVA. Ogni caso: informazioni sul cluster, distanza dal centro del cluster.

Considerazioni sui dati relativi all'analisi del cluster delle k-medie

Dati. Le variabili devono essere quantitative a livello di intervallo o di rapporto. Se le variabili sono binarie o conteggi, utilizzare la procedura Analisi cluster gerarchica.

Ordine dei casi e centri cluster iniziali. L'algoritmo predefinito per la scelta dei centri di cluster iniziali varia a seconda dell'ordine dei casi. L'opzione**Usa medie mobili** della finestra di dialogo Itera rende la soluzione risultante potenzialmente dipendente dall'ordine dei casi, indipendentemente dai centri di cluster scelti inizialmente. Se si utilizza uno di questi metodi, può essere utile ottenere più soluzioni diverse con casi disposti in ordini casuali diversi per verificare la stabilità di una soluzione specifica. Per evitare problemi con l'ordine dei casi, è consigliabile specificare i centri di cluster iniziali ed evitare di usare l'opzione **Usa medie mobili**. Tuttavia, l'ordinamento dei centri di cluster iniziali può influire sulla soluzione se esistono distanze assegnate dai casi ai centri di cluster. Per valutare la stabilità di una soluzione, è possibile confrontare i risultati delle analisi con diverse permutazioni dei valori dei centri iniziali.

Ipotesi. Le distanze vengono calcolate utilizzando la distanza euclidea semplice. Se si desidera utilizzare un'altra misura di distanza o di similarità, utilizzare la procedura Analisi cluster gerarchica. Lo scaling delle variabili è un'operazione che deve essere effettuata con molta attenzione. Se le variabili vengono misurate con scale diverse (ad esempio se una variabile è espressa in dollari e un'altra è espressa in anni), i risultati possono essere fuorvianti. In questi casi è consigliabile standardizzare le variabili prima di procedere con l'analisi cluster *k*medie (utilizzando la procedura Descrittive). Questa procedura presume che sia stato selezionato il numero esatto di cluster e che siano state incluse tutte le variabili rilevanti. Se è stato selezionato un numero di cluster inesatto o sono state omesse variabili importanti, i risultati possono essere inattendibili.

Per ottenere un'Analisi del cluster delle k-medie

1. Dai menu, scegliere:

   Analizza > **Classifica** > **Cluster delle k-medie...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare le variabili da utilizzare nell'analisi cluster.
3. Specificare il numero di cluster. Il numero di cluster specificato deve essere almeno di 2 e non deve essere maggiore al numero di casi del file di dati.
4. Selezionare il metodo **Itera e classifica** oppure il metodo **Classifica soltanto**.
5. In alternativa, selezionare una variabile di identificazione per etichettare i casi.

Efficienza dell'analisi del cluster delle k-medie

Il comando Cluster *k*-medie è efficace principalmente in quanto non calcola le distanze tra tutte le coppie di casi, a differenza di numerosi algoritmi di cluster, ad esempio quello utilizzato dal comando per la Cluster gerarchica di SPSS.

Per ottenere la massima efficienza, creare un campione di casi e utilizzare il metodo **Itera e classifica** per determinare i centri cluster. Selezionare **Scrivi valori finali su file**. Quindi, ripristinare tutto il file di dati e selezionare **Classifica soltanto** come metodo e selezionare **Leggi valori iniziali** per classificare tutto il file utilizzando i centri valutati per il campione. È possibile leggere o scrivere da un file o dataset. I dataset possono anche essere riutilizzati nella stessa sessione, ma non vengono salvati come file a meno che siano stati salvati come tali al termine della sessione. I nomi dei dataset devono essere conformi alle regole di denominazione delle variabili. Per ulteriori informazioni, consultare l'argomento .

Analisi del cluster delle k-medie: Itera

Nota: queste opzioni sono disponibili solo se si seleziona il metodo **Itera e classifica** dalla finestra di dialogo Analisi del cluster delle k-medie.

Numero massimo di iterazioni. Consente di impostare il numero massimo di iterazioni per l'algoritmo *k*-medie. Le iterazioni si interromperanno al numero impostato, anche se il criterio di convergenza non viene soddisfatto. Il numero deve essere compreso tra 1 e 999.

Per riprodurre l'algoritmo utilizzato dal comando Quick Cluster delle versioni di SPSS precedenti alla 5.0, impostare l'opzione **Massimo numero di iterazioni** su 1.

Criterio di convergenza. Determina il termine dell'iterazione. Rappresenta una percentuale della distanza minima tra i centri iniziali del cluster, quindi deve essere superiore a 0 ma non superiore a 1. Se il criterio è uguale a 0.02, ad esempio, l'iterazione cessa quando un'iterazione completa non sposta nessuno dei centri di cluster di una distanza superiore al 2% della distanza più piccola tra i centri di cluster iniziali.

Usa medie mobili. Consente di richiedere l'aggiornamento dei centri cluster in seguito all'assegnazione di ciascun caso Se non viene selezionata questa opzione, i nuovi centri del cluster verranno calcolati quando tutti i casi saranno stati assegnati.

Analisi del cluster delle k-medie: Salva

È possibile salvare informazioni sulla soluzione come nuove variabili da utilizzare in analisi successive:

Appartenenza cluster. Consente di creare una nuova variabile che indica l'appartenenza finale al cluster di ciascun caso. I valori della nuova variabile sono compresi tra 1 e il numero di cluster.

Distanza dal centro del cluster. Consente di creare una nuova variabile che indica la distanza euclidea tra ciascun caso e il relativo centro di classificazione.

Analisi del cluster delle k-medie: Opzioni

Statistiche. È possibile selezionare le seguenti statistiche: centri cluster iniziali, tabella ANOVA e informazioni sul cluster per ciascun caso.

- *Centri cluster iniziali*.Prima stima delle medie delle variabili per ciascuno dei cluster. In mancanza di indicazioni particolari, viene selezionato dai dati un numero casuale ben distanziati uguale al numero dei cluster. I centri dei cluster iniziali vengono usati per un primo ciclo di classificazione e poi vengono aggiornati.
- *Tabella ANOVA*. Visualizza una tabella di analisi della varianza che include test F univariati per ogni variabile di raggruppamento. I test F sono descrittivi e il livello di significatività fornisce informazioni utili. La tabella ANOVA non viene visualizzata se tutti i casi sono assegnati a un solo cluster.
- *Informazioni cluster per ogni caso*.Visualizza per ogni caso l'assegnazione finale del cluster e la distanza euclidea tra il caso e il centro del cluster utilizzata per classificarlo. Visualizza anche la distanza euclidea tra i centri del cluster finali.

Valori mancanti. Le opzioni disponibili sono **Escludi casi listwise** o **Escludi casi pairwise**.

- **Escludi casi listwise.** Consente di escludere i casi coni valori mancanti per le variabili di raggruppamento dall'analisi.
- **Escludi casi pairwise.** Consente di assegnare i casi ai cluster in base alle distanze calcolate da tutte le variabili con valori non mancanti.

Funzioni aggiuntive del comando QUICK CLUSTER

La Procedura K-Means Cluster utilizza la sintassi del comando `QUICK CLUSTER` . Il linguaggio della sintassi dei comandi consente inoltre di:

- Accettare i primi *k* casi come centri dei cluster iniziali per evitare il passaggio di dati normalmente usato per stimarli.
- Specificare i centri iniziali dei cluster direttamente come parte della sintassi del comando.
- Specificare i nomi delle variabili salvate.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test non parametrici

I test non parametrici formulano ipotesi minime sulla distribuzione sottostante dei dati. I test disponibili in queste finestre di dialogo si possono raggruppare in tre categorie generali a seconda di come sono organizzati i dati:

- Un test a campione singolo analizza un solo campo.
- Un test a campioni correlati confronta due o più campi per lo stesso insieme di casi.
- Un test di campioni indipendenti analizza un campo raggruppato secondo le categorie di un altro campo.

Test non parametrici a campione singolo

I test non parametrici a campione singolo identificano le differenze nei singoli campi utilizzando uno o più test non parametrici. I test non parametrici non presumono che i dati seguano la distribuzione normale.

Qual è il proprio l'obiettivo? Gli obiettivi consentono di specificare rapidamente varie impostazioni di uso comune per i test.

- **Confronta automaticamente i dati osservati con quelli ipotizzati.** Questo obiettivo applica il test binomiale ai campi categoriali con due sole categorie, il test del chi-quadrato a tutti gli altri campi categoriali e il test di Kolmogorov-Smirnov ai campi continui.
- **Prova la casualità della sequenza.** Questo obiettivo utilizza il test delle successioni per verificare la casualità della sequenza di valori dei dati osservata.
- **Analisi personalizzata.** Selezionare questa opzione se si desidera modificare manualmente le impostazioni del test nella scheda Impostazioni. Si osservi che questa impostazione viene selezionata automaticamente se successivamente si apportano delle modifiche alle opzioni nella scheda Impostazioni che sono incompatibili con l'obiettivo attualmente selezionato.

Ottenimento di One - Sample Nonparametric Test

Dai menu, scegliere:

Analizza > **Test non parametrici** > **Campione singolo...**

1. Fare clic su **Esegui**.

Facoltativamente, è possibile:

- Specificare un obiettivo nella scheda Obiettivi.
- Specificare le assegnazioni di campo nella scheda Campi.
- Specificare delle impostazioni avanzate nella scheda Impostazioni.

Scheda Campi

La scheda Campi indica i campi che è necessario testare.

Utilizza ruoli predefiniti. Questa opzione utilizza le informazioni contenute nei campi esistenti. Tutti i campi con un ruolo predefinito come Input, Obiettivo o Entrambi saranno utilizzati come campi test. È obbligatorio avere almeno un campo test.

Utilizza assegnazioni campi personalizzate. Questa opzione consente di ignorare i ruoli dei campi. Dopo averla selezionata, compilare i campi riportati sotto:

- **Campi test.** Selezionare uno o più campi.

Scheda Impostazioni

La scheda Impostazioni contiene vari gruppi di impostazioni differenti che è possibile modificare per definire il modo in cui l'algoritmo elabora i dati. Se si apportano delle modifiche alle impostazioni predefinite che sono incompatibili con l'obiettivo attualmente selezionato, la scheda Obiettivo viene aggiornata automaticamente per selezionare l'opzione **Personalizza analisi**.

Scegli test

Queste impostazioni specificano i test da eseguire sui campi indicati nella scheda Campi.

Scegli automaticamente i test in base ai dati. Questa impostazione applica il test binomiale ai campi categoriali con due sole categorie valide (non mancanti), il test del chi-quadrato a tutti gli altri campi categoriali e il test di Kolmogorov-Smirnov ai campi continui.

Personalizza i test. Questa impostazione consente di definire l'esecuzione di test specifici.

- **Confronta la probabilità binaria osservata con quella ipotizzata (test binomiale).** Il test binomiale può essere applicato a tutti i campi. Esso genera un test a campione singolo che verifica se la distribuzione osservata di un campo indicatore (campo categoriale con due sole categorie) è uguale a quella prevista da una distribuzione binomiale specificata. È possibile inoltre richiedere gli intervalli di confidenza. Consultare “Test binomiale: Opzioni ” a pagina 234 per i dettagli sulle impostazioni del test.
- **Confronta le probabilità osservate con quelle ipotizzate (test chi-quadrato).** Il test del chi-quadrato viene applicato ai campi nominali e ordinali. Questa opzione genera un test a campione singolo che calcola una statistica chi-quadrato in base alle differenze tra le frequenze osservate e previste delle categorie di un campo. Consultare “Test del chi-quadrato: Opzioni ” a pagina 235 per i dettagli sulle impostazioni del test.
- **Prova la distribuzione osservata con quella ipotizzata (test di Kolmogorov-Smirnov).** Il test di Kolmogorov-Smirnov viene applicato ai campi continui e ordinali. Questa opzione genera un test a campione singolo che verifica se la funzione di distribuzione cumulativa del campione di un campo è omogenea con una distribuzione uniforme, normale, di Poisson o esponenziale. Consultare “Opzioni test di Kolmogorov-Smirnov ” a pagina 235 per i dettagli sulle impostazioni del test.
- **Confronta mediana osservata con quella ipotizzata (test dei ranghi con segno di Wilcoxon).** Il test dei ranghi con segno di Wilcoxon si applica ai campi continui e ordinali. Questa opzione genera un test a campione singolo del valore mediano di un campo. Specificare un numero come mediana ipotizzata.
- **Prova la casualità della sequenza (test delle successioni).** Il test delle successioni viene applicato a tutti i campi. Questa opzione genera un test a campione singolo che verifica se la sequenza dei valori di un campo dicotomizzato è casuale. Consultare “Opzioni test delle successioni ” a pagina 235 per i dettagli sulle impostazioni del test.

Test binomiale: Opzioni

Il test binomiale è destinato ai campi indicatore (campi categoriali con due sole categorie), ma viene applicato a tutti i campi utilizzando le regole per la definizione dell'"esito positivo".

Proporzione ipotizzata. Questo specifica la proporzione prevista di record definiti come "successi", o *p*. Specificare un valore superiore a 0 e inferiore a 1. Il valore predefinito è 0,5.

Intervallo di confidenza. Sono disponibili i seguenti metodi per calcolare gli intervalli di confidenza per i dati binari.

- **Clopper-Pearson (esatto).** Un intervallo esatto basato sulla distribuzione binomiale cumulativa.
- **Jeffreys.** Un intervallo bayesiano basato sulla distribuzione a posteriori di *p* utilizzando la probabilità a priori di Jeffreys.
- **Rapporto di verosimiglianza.** Un intervallo basato sulla funzione di verosimiglianza per *p*.

Definisci l'esito positivo per i campi categoriali. Specifica come viene definito l'"esito positivo" (il valore o i valori dei dati confrontati con la proporzione ipotizzata) per i campi categoriali.

- **Utilizza la prima categoria trovata nei dati** esegue il test binomiale utilizzando il primo valore trovato nel campione per definire l'"esito positivo". Questa opzione si può applicare solo ai campi nominali o ordinali con due soli valori; tutti gli altri campi categoriali specificati nella scheda Campi in cui è utilizzata questa opzione non vengono testati. È l'impostazione predefinita.
- **Specifica valori dell'esito positivo** esegue il test binomiale utilizzando l'elenco dei valori specificati per definire l'"esito positivo". Specificare un elenco di valori stringa o numerici. I valori dell'elenco non devono necessariamente essere presenti nel campione.

Definisci l'esito positivo per i campi continui. Specifica come viene definito l'"esito positivo" (il valore o i valori dei dati confrontati con il valore del test) per i campi continui. L'esito positivo viene definito come valori uguali o minori di un punto di divisione.

- **Valore intermedio campione** imposta il punto di divisione sulla media dei valori massimo e minimo.
- **Punto di divisione personalizzato** consente di specificare un valore per il punto di divisione.

Test del chi-quadrato: Opzioni

Tutte le categorie hanno uguale probabilità. Genera frequenze uguali fra tutte le categorie del campione. È l'impostazione predefinita.

Personalizza probabilità prevista. Consente di indicare frequenze non uguali per un elenco specificato di categorie. Specificare un elenco di valori stringa o numerici. I valori dell'elenco non devono necessariamente essere presenti nel campione. Nella colonna **Categoria**, specificare i valori delle categorie. Nella colonna **Frequenza relativa**, specificare un valore maggiore di 0 per ogni categoria. Le frequenze personalizzate vengono considerate rapporti: così, ad esempio, specificare le frequenze 1, 2 e 3 equivale a specificare le frequenze 10, 20 e 30, ed entrambe specificano che si presume che 1/6 dei record ricada nella prima categoria, 1/3 nella seconda e 1/2 nella terza. Quando si specificano le probabilità previste personalizzate, i valori delle categorie personalizzate devono includere tutti i valori dei campi nei dati; altrimenti, il test non viene eseguito per quel campo.

Opzioni test di Kolmogorov-Smirnov

Questa finestra di dialogo specifica le distribuzioni da testare e i parametri delle distribuzioni ipotizzate.

Quando alcuni parametri della distribuzione devono essere stimati dal campione, il test Kolmogorov-Smirnov non si applica più. In queste istanze, la statistica di prova di Lilliefors può essere utilizzata per stimare il valore *p*utilizzando il campionamento di Monte Carlo per la verifica della normalità con media e varianza sconosciuta. Il test di Lilliefors si applica alle tre distribuzioni continue (**Normal**, **Exponentia**le **Uniform**). Si noti che il test non si applica se la distribuzione sottostante è discreta (**Poisson**). Il test è definito solo per l'inferenza di un campione quando i parametri di distribuzione corrispondenti non sono specificati.

Normale
: **Utilizza i dati di esempio** utilizza la deviazione media e standard osservata e fornisce le opzioni per la selezione dei risultati esistenti di **test Asintomtotic** oppure utilizzare **Lilliefors test basato sul campionamento di Monte Carlo**. **Custom** consente di specificare i valori.

Uniforme
: **Utilizza i dati di esempio** utilizza il minimo e il massimo osservato e utilizza il test di Lilliefors basato sul campionamento di Monte Carlo. **Custom** consente di specificare valori minimi e massimi.

Esponenziale
: **media di esempio** utilizza la media osservata e utilizza il test di Lilliefors basato sul campionamento di Monte Carlo. **Custom** consente di specificare un valore medio osservato.

Poisson
: **Media** consente di specificare un valore medio osservato.

Opzioni test delle successioni

Il test delle successioni è destinato ai campi indicatore (campi categoriali con due sole categorie), ma può essere applicato a tutti i campi utilizzando le regole per la definizione dei gruppi.

Definisci i gruppi per i campi categoriali. Sono disponibili le seguenti opzioni:

- **Il campione contiene solo 2 categorie** esegue il test delle successioni definendo i gruppi con i valori rilevati nel campione. Questa opzione si può applicare solo ai campi nominali o ordinali con due soli valori; tutti gli altri campi categoriali specificati nella scheda Campi in cui è utilizzata questa opzione non vengono testati.
- **Ricodifica i dati in 2 categorie** esegue il test delle successioni definendo uno dei gruppi con l'elenco di valori specificato. Tutti gli altri valori del campione definiscono l'altro gruppo. Non è necessario che tutti i valori dell'elenco siano presenti nel campione, ma ogni gruppo deve comprendere almeno un record.

Definisci il punto di divisione per i campi continui. Specifica come vengono definiti i gruppi per i campi continui. Il primo gruppo è definito come valori uguali o inferiori a un punto di divisione.

- **Mediana campione** imposta il punto di divisione sulla mediana del campione.
- **Media del campione** imposta il punto di divisione sulla media del campione.
- **Personalizzato** consente di specificare un valore per il punto di divisione.

Opzioni test

Livello di significatività
Specifica il livello di significatività (alfa) di tutti i test. Indicare un valore numerico compreso fra 0 e 1. L'impostazione predefinita è 0,05.

Intervallo di confidenza (%)
Specifica il livello di confidenza per tutti gli intervalli di confidenza generati. Specificare un valore numerico compreso tra 0 e 100. 95 è il default.

Casi esclusi
Specifica come determinare la base di casi per i test.

Escludi i casi test per test
I record con valori mancanti per un campo che viene utilizzato per un test specifico sono omessi da quel test. Quando nell'analisi vengono specificati più test, ciascuno viene valutato separatamente.

Escludi casi listwise
I record con valori mancanti per qualsiasi campo che si chiama nella scheda Campi sono esclusi da tutte le analisi.

Campionamento di Monte Carlo

Quando alcuni parametri della distribuzione devono essere stimati dal campione, il test Kolmogorov-Smirnov non si applica più. In queste istanze, la statistica di prova di Lilliefors può essere utilizzata per stimare il valore *p*utilizzando il campionamento di Monte Carlo per la verifica della normalità con media e varianza sconosciuta. Il test di Lilliefors si applica alle tre distribuzioni continue (**Normal**, **Exponentia**l**e Uniform**). Si noti che il test non si applica se la distribuzione sottostante è discreta (**Poisson**). Il test è definito solo per l'inferenza di un campione quando i parametri di distribuzione corrispondenti non sono specificati.

Imposta seme personalizzato
Quando è abilitata, questa impostazione fornisce l'opzione di reimpostazione del valore casuale **Seed** utilizzato per il campionamento di Monte Carlo. Il valore deve essere un intero intero compreso tra 1 e 2.147.483.647 (il valore predefinito è 2.000.000). Per impostazione predefinita, l'impostazione non è abilitata, che produce un valore seed casuale.

Numero di campioni
Reimposta il numero di repliche di campionamento di Monte Carlo che vengono utilizzati dal test di Lilliefors. Il valore deve essere un intero intero compreso tra 100 e il numero intero più grande. Il valore predefinito è 10.000.

Livello di confidenza di simulazione (%)
Reimposta il livello di intervallo di confidenza stimato del test di Kolmogorov-Smirnov. Il valore deve essere un valore unico compreso tra 0 e 100. Il valore predefinito è 99.

Valori mancanti definiti dall'utente

Valori mancanti definiti dall'utente per i campi categoriali. Per essere inclusi nell'analisi, i campi categoriali devono contenere valori validi per un record. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente devono essere considerati validi nell'ambito dei campi categoriali. I valori mancanti di sistema e i valori mancanti relativi ai campi continui vengono sempre considerati non validi.

Funzioni aggiuntive del comando NPTESTS

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare test a campione singolo, a campioni indipendenti e a campioni correlati in un'unica esecuzione della procedura.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test di Wilcoxon a campione singolo

Il test dei ranghi con segno di Wilcoxon è un test statistico non parametrico utilizzato per verificare la posizione di una popolazione in base ai dati campione o per confrontare le posizioni di due popolazioni. Il test dei ranghi con segno di Wilcoxon a un campione è un'alternativa non parametrica al test t del campione quando i dati non seguono una distribuzione normale. Consente di determinare se esiste una differenza significativa tra la mediana di un campione e un valore mediano ipotizzato. Qui la statistica del test viene calcolata in base ai ranghi della differenza tra i valori osservati e il valore ipotizzato.

Il test dei ranghi con segno di Wilcoxon assume che la popolazione dei genitori sia simmetrica rispetto alla mediana. Poiché i campioni casuali provengono da una popolazione con una funzione di distribuzione cumulativa continua e sono simmetrici rispetto alla mediana, sotto l'ipotesi nulla, le differenze tra i valori osservati e il valore ipotizzato sono simmetricamente distribuite intorno allo zero in modo che le differenze positive e negative di uguale grandezza assoluta abbiano la stessa probabilità di occorrenza.

Esempio

Supponiamo che una nuova società di startup sia interessata ad assumere ingegneri software, ma vuole assicurarsi che stiano offrendo stipendi competitivi. Si supponga che abbiano raccolto dati salariali da un campione di 30 ingegneri software attualmente impiegati in varie aziende dello stesso settore. Se la società desidera stabilire se lo stipendio mediano che stanno offrendo ai loro potenziali assunti è significativamente diverso dallo stipendio mediano degli ingegneri del software nel settore, è possibile utilizzare un rango di segno Wilcoxon di esempio per questo scopo. In questo caso, si assume l'ipotesi null che lo stipendio mediano offerto dalla startup non sia significativamente diverso dallo stipendio mediano nel settore, ovvero che la differenza di stipendio mediana sia zero.

Per eseguire il test Wilcoxon di un campione

(Questa funzione richiede l'opzione Statistics Base.)

Dal menu scegliere:

1. **Analisi> Test non parametrici> Un campione ...**
2. Fare clic su **'Impostazioni'**.
3. Sotto **'Seleziona un elemento'**, selezionare **'Scegli test'**.
4. Scegliere **'Personalizza test'**.
5. Selezionare **'Confronta mediana con quella ipotizzata (test dei ranghi con segno di Wilcoxon)'**.
6. Immettere il valore della **'Mediana ipotizzata'** e fare clic su **' Esegui**.

Per ulteriori informazioni, vedere https://www.ibm.com/support/pages/can-spss-statistics-produce-one-sample-wilcoxon-test

Test non parametrici di campioni indipendenti

I test non parametrici di campioni indipendenti individuano le differenze fra due o più gruppi mediante uno o più test non parametrici. I test non parametrici non presumono che i dati seguano la distribuzione normale.

Qual è il proprio l'obiettivo? Gli obiettivi consentono di specificare rapidamente varie impostazioni di uso comune per i test.

- **Confronta automaticamente le distribuzioni nei gruppi.** Questo obiettivo applica il test U di Mann-Whitney ai dati con 2 gruppi o il test ANOVA a una via di Kruskal-Wallis ai dati con *k* gruppi.
- **Confronta le mediane nei gruppi.** Questo obiettivo utilizza il test della mediana per confrontare le mediane osservate tra più gruppi.

- **Analisi personalizzata.** Selezionare questa opzione se si desidera modificare manualmente le impostazioni del test nella scheda Impostazioni. Si osservi che questa impostazione viene selezionata automaticamente se successivamente si apportano delle modifiche alle opzioni nella scheda Impostazioni che sono incompatibili con l'obiettivo attualmente selezionato.

Per ottenere test non parametrici di campioni indipendenti

Dai menu, scegliere:

Analizza > **Test non parametrici** > **Campioni indipendenti...**

1. Fare clic su **Esegui**.

Facoltativamente, è possibile:

- Specificare un obiettivo nella scheda Obiettivi.
- Specificare le assegnazioni di campo nella scheda Campi.
- Specificare delle impostazioni avanzate nella scheda Impostazioni.

Scheda Campi

La scheda Campi indica i campi da testare e il campo utilizzato per definire i gruppi.

Utilizza ruoli predefiniti. Questa opzione utilizza le informazioni contenute nei campi esistenti. Tutti i campi continui e ordinali con un ruolo predefinito come Obiettivo o Entrambi saranno utilizzati come campi test. Se vi è un solo campo categoriale con un ruolo predefinito come Input, verrà utilizzato come un campo di raggruppamento. Altrimenti, per impostazione predefinita non viene utilizzato alcun campo di raggruppamento ed è necessario utilizzare le assegnazioni campi personalizzate. È obbligatorio disporre di almeno un campo test e di un campo di raggruppamento.

Utilizza assegnazioni campi personalizzate. Questa opzione consente di ignorare i ruoli dei campi. Dopo averla selezionata, compilare i campi riportati sotto:

- **Campi test.** Selezionare uno o più campi continui od ordinali.
- **Gruppi.** Selezionare un campo categoriale.

Scheda Impostazioni

La scheda Impostazioni contiene vari gruppi di impostazioni che è possibile modificare per perfezionare l'elaborazione dei dati da parte dell'algoritmo. Se si apportano delle modifiche alle impostazioni predefinite che sono incompatibili con l'obiettivo attualmente selezionato, la scheda Obiettivo viene aggiornata automaticamente per selezionare l'opzione **Personalizza analisi**.

Scegli test

Queste impostazioni specificano i test da eseguire sui campi indicati nella scheda Campi.

Scegli automaticamente i test in base ai dati. Questa impostazione applica il test U di Mann-Whitney ai dati con 2 gruppi o il test ANOVA a una via di Kruskal-Wallis ai dati con *k* gruppi.

Personalizza i test. Questa impostazione consente di definire l'esecuzione di test specifici.

- **Confronta le distribuzioni nei gruppi.** Questa impostazione genera test di campioni indipendenti per verificare se i campioni appartengono alla stessa popolazione.

  U di Mann-Whitney (2 campioni) utilizza il rango di ogni caso per verificare se i gruppi sono estratti dalla stessa popolazione. Il primo valore in ordine crescente del campo di raggruppamento definisce il primo gruppo, mentre il secondo definisce il secondo gruppo. Se il campo di raggruppamento ha più di due valori, il test non viene generato.

  Kolmogorov-Smirnov (2 campioni) è sensibile a tutte le differenze di mediana, dispersione, asimmetria e simili fra le due distribuzioni. Se il campo di raggruppamento ha più di due valori, il test non viene generato.

Sequenza di test per casualità (Wald-Wolfowitz per 2 campioni) genera un test delle successioni secondo il criterio dell'appartenenza a un gruppo. Se il campo di raggruppamento ha più di due valori, il test non viene generato.

ANOVA a 1 via di Kruskal-Wallis (k campioni) è un'ampliamento del test U di Mann-Whitney ed è l'analogo non parametrico dell'analisi della varianza a una via. Se lo si desidera è possibile richiedere confronti multipli dei *k* campioni, ovvero confronti multipli **tutto pairwise** o confronti **decrescenti stepwise**.

Test per alternative ordinate (Jonckheere-Terpstra per k campioni) è un'alternativa più potente al test di Kruskal-Wallis quando i *k* campioni hanno un ordinamento naturale. Ad esempio, le *k* popolazioni possono rappresentare *k* temperature crescenti. L'ipotesi che diverse temperature producano la stessa distribuzione della risposta è verificata rispetto all'ipotesi alternativa in base a cui al salire della temperatura, cresce la grandezza della risposta. Qui l'ipotesi alternativa è ordinata e quindi il test di Jonckheere-Terpstra è il più appropriato da utilizzare. **Dal più piccolo al più grande** specifica l'ipotesi alternativa che il parametro ubicazione del primo gruppo è inferiore o uguale al secondo, che a sua volta è inferiore o uguale al terzo, e così via. **Dal più grande al più piccolo** specifica l'ipotesi alternativa che il parametro ubicazione del primo gruppo è superiore o uguale al secondo, che a sua volta è superiore o uguale al terzo, e così via. Per entrambe le opzioni, l'ipotesi alternativa presume anche che le ubicazioni non siano tutte uguali. Se lo si desidera, è possibile richiedere confronti multipli dei *k* campioni, ovvero confronti multipli di tipo **Tutto pairwise** o confronti di tipo **Decrescente stepwise**.

- **Confronta gli intervalli nei gruppi.** Questa opzione genera un test di campioni indipendenti per verificare se i campioni hanno lo stesso intervallo. **Reazioni estreme di Moses (2 campioni)** verifica un gruppo di controllo con un gruppo di confronto. Il primo valore in ordine crescente del campo di raggruppamento definisce il gruppo di controllo, mentre il secondo definisce il gruppo di confronto. Se il campo di raggruppamento ha più di due valori, il test non viene generato.
- **Confronta le mediane nei gruppi.** Questa opzione genera un test di campioni indipendenti per verificare se i campioni hanno la stessa mediana. **Test della mediana (k campioni)** può utilizzare come mediana ipotizzata sia la mediana campione raggruppata (calcolata su tutti i record del dataset), sia un valore personalizzato. Se lo si desidera, è possibile richiedere confronti multipli dei *k* campioni, ovvero confronti multipli di tipo **Tutto pairwise** o confronti di tipo **Decrescente stepwise**.
- **Stima l'intervallo di confidenza nei gruppi. Stima di Hodges-Lehman (2 campioni)** genera una stima a campioni indipendenti e un intervallo di confidenza per la differenza tra le mediane di due gruppi. Se il campo di raggruppamento ha più di due valori, il test non viene generato.

Opzioni test

Livello di significatività. Specifica il livello di significatività (alfa) di tutti i test. Indicare un valore numerico compreso fra 0 e 1. L'impostazione predefinita è 0,05.

Intervallo di confidenza (%). Questo specifica il livello di confidenza per tutti gli intervalli di confidenza prodotti. Specificare un valore numerico compreso tra 0 e 100. 95 è il default.

Casi esclusi. Specifica come determinare la base di casi per i test. **Escludi casi listwise** significa che i record con valori mancanti in qualunque campo denominato in un sottocomando vengono esclusi da tutte le analisi. **Escludi casi test per test** significa che i record con valori mancanti per un campo utilizzato in un determinato test vengono omessi da quel test. Quando nell'analisi vengono specificati più test, ciascuno viene valutato separatamente.

Valori mancanti definiti dall'utente

Valori mancanti definiti dall'utente per i campi categoriali. Per essere inclusi nell'analisi, i campi categoriali devono contenere valori validi per un record. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente devono essere considerati validi nell'ambito dei campi categoriali. I valori mancanti di sistema e i valori mancanti relativi ai campi continui vengono sempre considerati non validi.

Funzioni aggiuntive del comando NPTESTS

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare test a campione singolo, a campioni indipendenti e a campioni correlati in un'unica esecuzione della procedura.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test non parametrici di campioni correlati

Identifica le differenze tra due o più campi correlati utilizzando uno o più test non parametrici. I test non parametrici non presumono che i dati seguano la distribuzione normale.

Considerazioni sui dati. Ogni record corrisponde a un determinato soggetto per il quale due o più misure correlate vengono memorizzate in campi separati del dataset. Ad esempio, uno studio relativo all'efficacia di un piano di dieta può essere analizzato utilizzando i test non parametrici a campioni correlati, se il peso di ciascun soggetto viene misurato a intervalli regolari e memorizzato in campi quali *Peso prima della dieta*, *Peso intermedio* e *Peso dopo la dieta*. Questi campi sono "correlati".

Qual è il proprio l'obiettivo? Gli obiettivi consentono di specificare rapidamente varie impostazioni di uso comune per i test.

- **Confronta automaticamente i dati osservati con quelli ipotizzati.** Questo obiettivo applica ai dati categoriali il test di McNemar quando vengono specificati 2 campi e il test Q di Cochran quando vengono specificati più di 2 campi; ai dati continui, il test dei ranghi con segno per confronti tra coppie di Wilcoxon quando vengono specificati 2 campi e il test ANOVA per ranghi a due vie di Friedman quando vengono specificati più di 2 campi.
- **Analisi personalizzata.** Selezionare questa opzione se si desidera modificare manualmente le impostazioni del test nella scheda Impostazioni. Si osservi che questa impostazione viene selezionata automaticamente se successivamente si apportano delle modifiche alle opzioni nella scheda Impostazioni che sono incompatibili con l'obiettivo attualmente selezionato.

Quando si specificano campi con livelli di misurazione diversi, essi vengono prima separati dal livello di misurazione, quindi si applica il test appropriato a ciascun gruppo. Ad esempio, se si sceglie **Confronta automaticamente i dati osservati con quelli ipotizzati** come obiettivo e si specificano 3 campi continui e 2 campi nominali, il test di Friedman viene applicato ai campi continui e il test di McNemar viene applicato ai campi nominali.

Per ottenere test non parametrici di campioni correlati

Dai menu, scegliere:

Analizza > **Test non parametrici** > **Campioni correlati...**

1. Fare clic su **Esegui**.

Facoltativamente, è possibile:

- Specificare un obiettivo nella scheda Obiettivi.
- Specificare le assegnazioni di campo nella scheda Campi.
- Specificare delle impostazioni avanzate nella scheda Impostazioni.

Scheda Campi

La scheda Campi indica i campi che è necessario testare.

Utilizza ruoli predefiniti. Questa opzione utilizza le informazioni contenute nei campi esistenti. Tutti i campi con un ruolo predefinito come Obiettivo o Entrambi saranno utilizzati come campi test. È obbligatorio avere almeno due campi test.

Utilizza assegnazioni campi personalizzate. Questa opzione consente di ignorare i ruoli dei campi. Dopo averla selezionata, compilare i campi riportati sotto:

- **Campi test.** Selezionare due o più campi. Ogni campo rappresenta un campione correlato diverso.

Scheda Impostazioni

La scheda Impostazioni contiene vari gruppi di impostazioni che è possibile modificare per perfezionare il modo in cui la procedura elabora i dati. Se si apportano modifiche alle impostazioni predefinite che risultano incompatibili con gli altri obiettivi, la scheda Obiettivo viene aggiornata automaticamente per selezionare l'opzione **Personalizza analisi**.

Scegli test

Queste impostazioni specificano i test da eseguire sui campi indicati nella scheda Campi.

Scegli automaticamente i test in base ai dati. Questa impostazione applica ai dati categoriali il test di McNemar quando vengono specificati 2 campi e il test Q di Cochran quando vengono specificati più di 2 campi; ai dati continui, il test dei ranghi con segno per confronti tra coppie di Wilcoxon quando vengono specificati 2 campi e il test ANOVA per ranghi a due vie di Friedman quando vengono specificati più di 2 campi.

Personalizza i test. Questa impostazione consente di definire l'esecuzione di test specifici.

- **Test per la modifica dei dati binari. McNemar's (2 campioni)** può essere applicato a campi categorici. Questa opzione genera un test a campioni correlati per verificare se le combinazioni di valori tra due campi indicatore (campi categoriali con due soli valori) hanno uguale probabilità. Se nella scheda Campi sono specificati più di due campi, il test non viene eseguito. Consultare “Test di McNemar: definisci esito positivo ” a pagina 241 per i dettagli sulle impostazioni del test. **Q di Cochran (k campioni)** si può applicare ai campi categoriali. Questa opzione genera un test a campioni correlati per verificare se le combinazioni di valori tra *k* campi indicatore (campi categoriali con due soli valori) hanno uguale probabilità. Se lo si desidera è possibile richiedere confronti multipli dei *k* campioni, ovvero confronti multipli **tutto pairwise** o confronti **decrescenti stepwise**. Consultare “Q di Cochran: definisci esito positivo ” a pagina 242 per i dettagli sulle impostazioni del test.
- **Test per le modifiche dei Dati Multinomiali. Test di omogeneità marginale (2 campioni)** produce un test di campioni relativi a se combinazioni di valori tra due campi ordinali accoppiati sono ugualmente probabili. Il test di omogeneità marginale si utilizza in genere nelle situazioni in cui sono presenti misure ripetute. Estensione del test di McNemar dalla risposta binaria a quella multinomiale. Se nella scheda Campi sono specificati più di due campi, il test non viene eseguito.
- **Confronta differenza mediana con quella ipotizzata.** Questi test producono ciascuno un test a campione correlato di se la differenza mediana tra due campi sia differente da 0. Il test si applica ai campi continui ed ordinali. Se nella scheda Campi sono specificati più di due campi, questi test non vengono eseguiti.
- **Stima intervallo di confidenza.** Genera una stima a campioni correlati e un intervallo di confidenza per la differenza mediana fra due campi accoppiati. Il test si applica ai campi continui ed ordinali. Se nella scheda Campi sono specificati più di due campi, questo test non viene eseguito.
- **Quantifica associazioni. Coefficiente di concordanza di Kendall (k campioni)** genera una misura di accordo tra giudici o stimatori in cui ogni record rappresenta il punteggio di vari elementi (campi) da parte di un giudice. Se lo si desidera, è possibile richiedere confronti multipli dei *k* campioni, ovvero confronti multipli di tipo **Tutto pairwise** o confronti di tipo **Decrescente stepwise**.
- **Confronta Distribuzioni. L'ANOVA a 2 vie di Friedman da parte dei ranghi (k campioni)** produce un test relativo ai campioni correlati se i campioni relativi a *k* sono stati tratti dalla stessa popolazione. Se lo si desidera, è possibile richiedere confronti multipli dei *k* campioni, ovvero confronti multipli di tipo **Tutto pairwise** o confronti di tipo **Decrescente stepwise**.

Test di McNemar: definisci esito positivo

Il test di McNemar è destinato ai campi indicatore (campi categoriali con due sole categorie), ma viene applicato a tutti i campi categoriali utilizzando le regole per la definizione dell'"esito positivo".

Definisci l'esito positivo per i campi categoriali. Specifica come viene definito l'"esito positivo" per i campi categoriali.

- **Utilizza la prima categoria trovata nei dati** esegue il test utilizzando il primo valore trovato nel campione per definire l'"esito positivo". Questa opzione si può applicare solo ai campi nominali o

ordinali con due soli valori; tutti gli altri campi categoriali specificati nella scheda Campi in cui è utilizzata questa opzione non vengono testati. È l'impostazione predefinita.

- **Specifica valori dell'esito positivo** esegue il test utilizzando l'elenco dei valori specificati per definire l'"esito positivo". Specificare un elenco di valori stringa o numerici. I valori dell'elenco non devono necessariamente essere presenti nel campione.

Q di Cochran: definisci esito positivo

Il test Q di Cochran è destinato ai campi indicatore (campi categoriali con due sole categorie), ma viene applicato a tutti i campi categoriali utilizzando le regole per la definizione dell'"esito positivo".

Definisci l'esito positivo per i campi categoriali. Specifica come viene definito l'"esito positivo" per i campi categoriali.

- **Utilizza la prima categoria trovata nei dati** esegue il test utilizzando il primo valore trovato nel campione per definire l'"esito positivo". Questa opzione si può applicare solo ai campi nominali o ordinali con due soli valori; tutti gli altri campi categoriali specificati nella scheda Campi in cui è utilizzata questa opzione non vengono testati. È l'impostazione predefinita.
- **Specifica valori dell'esito positivo** esegue il test utilizzando l'elenco dei valori specificati per definire l'"esito positivo". Specificare un elenco di valori stringa o numerici. I valori dell'elenco non devono necessariamente essere presenti nel campione.

Opzioni test

Livello di significatività. Specifica il livello di significatività (alfa) di tutti i test. Indicare un valore numerico compreso fra 0 e 1. L'impostazione predefinita è 0,05.

Intervallo di confidenza (%). Questo specifica il livello di confidenza per tutti gli intervalli di confidenza prodotti. Specificare un valore numerico compreso tra 0 e 100. 95 è il default.

Casi esclusi. Specifica come determinare la base di casi per i test.

- **Escludi casi listwise** significa che i record con valori mancanti in qualunque campo denominato in un sottocomando vengono esclusi da tutte le analisi.
- **Escludi casi test per test** significa che i record con valori mancanti per un campo utilizzato in un determinato test vengono omessi da quel test. Quando nell'analisi vengono specificati più test, ciascuno viene valutato separatamente.

Valori mancanti definiti dall'utente

Valori mancanti definiti dall'utente per i campi categoriali. Per essere inclusi nell'analisi, i campi categoriali devono contenere valori validi per un record. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente devono essere considerati validi nell'ambito dei campi categoriali. I valori mancanti di sistema e i valori mancanti relativi ai campi continui vengono sempre considerati non validi.

Funzioni aggiuntive del comando NPTESTS

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare test a campione singolo, a campioni indipendenti e a campioni correlati in un'unica esecuzione della procedura.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Funzioni aggiuntive del comando NPTESTS

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare test a campione singolo, a campioni indipendenti e a campioni correlati in un'unica esecuzione della procedura.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Finestre legacy

Esistono anche numerose finestre di dialogo "legacy" che eseguono test non parametrici. Queste finestre di dialogo supportano la funzionalità fornita dal modulo Exact Tests.

Test del chi-quadrato. Consente di ordinare in tabelle una variabile in categorie e di ottenere una statistica chi-quadrato in base alle differenze tra frequenze osservate e attese.

Test binomiale. Consente di confrontare la frequenza osservata in ciascuna categoria di una variabile dicotomica con le frequenze attese dalla distribuzione binomiale.

Test delle successioni. Consente di verificare se l'ordine di occorrenza di due valori di una variabile è casuale.

Test di Kolmogorov-Smirnov per un campione. Consente di confrontare la funzione di distribuzione cumulativa osservata per una variabile con la distribuzione teorica specificata, che può essere normale, uniforme, esponenziale o di Poisson.

Test di due campioni indipendenti. Consente di confrontare due gruppi di casi in base a una sola variabile. Sono disponibili il test *U* di Mann-Whitney, il test di Kolmogorov-Smirnov per due campioni, il test delle reazioni estreme di Moses e il test delle successioni di Wald-Wolfowitz.

Test per due campioni correlati. Consente di confrontare le distribuzioni di due variabili. Sono disponibili il test dei ranghi con segno di Wilcoxon, il test dei segni e il test di McNemar.

Test per diversi campioni indipendenti. Consente di confrontare due o più gruppi di casi in base alla stessa variabile. Sono disponibili il test di Kruskal-Wallis, il test della mediana e il test di Jonckheere-Terpstra.

Test per diversi campioni correlati. Consente di confrontare le distribuzioni di due o più variabili. Sono disponibili il test di Friedman, il test *W* di Kendall e il test *Q* di Cochran.

Per tutti i test precedentemente citati sono disponibili quartili e media, deviazione standard, valore minimo e massimo e numero di casi non mancanti.

Test del chi-quadrato

La procedura Test del chi-quadrato permette di ordinare in tabelle una variabile in categorie e di calcolare una statistica chi-quadrato. Questo test sulla bontà di adattamento permette di confrontare le frequenze osservate e attese in ciascuna categoria per verificare se tutte le categorie includono la stessa proporzione di valori o se includono una proporzione di valori specificati dall'utente.

Esempi. Il test del chi-quadrato può essere utilizzato per determinare se in un sacchetto di gelatine di frutta è presente la stessa proporzione di blu, marrone, arancio, rosso e giallo. È inoltre possibile determinare se il sacchetto di gelatine contiene il 5% di blu, il 30% di marrone, il 10% di verde, il 20% di arancio, il 15% di rosso e il 15% di giallo.

Statistiche. Media, deviazione standard, valore minimo e massimo e quartili. Il numero e la percentuale di casi mancanti e non mancanti, il numero di casi osservati e attesi per ciascuna categoria, i residui e la statistica chi-quadrato.

Considerazioni sui dati di test del chi-quadrato

Dati. Utilizzare variabili categoriali numeriche ordinate o non ordinate (livelli di misurazione ordinali o nominali). Per convertire le variabili stringa in variabili numeriche, utilizzare il comando Ricodifica automatica del menu Trasforma.

Ipotesi. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Si presume che i dati rappresentino un campione casuale. Le frequenze previste per ogni categoria dovrebbero essere almeno 1. Non più del 20% delle categorie avrebbe dovuto prevedere frequenze inferiori a 5.

Per ottenere un test del chi-quadrato

1. Dai menu, scegliere:

Analizza > **Test non parametrici** > **Finestre di dialogo legacy** > **Chi-quadrato...**

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una o più variabili per il test. Ogni variabile produce un test distinto.
3. È possibile fare clic su **Opzioni** per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

Intervallo e valori attesi del test del chi-quadrato

Intervallo atteso. Per impostazione predefinita, ogni singolo valore della variabile è definito come categoria. Per definire categorie all'interno di un intervallo specifico, selezionare **Usa intervallo specificato** e inserire valori interi per il limite inferiore e superiore. Le categorie verranno definite per ogni valore intero incluso nell'intervallo specificato. I casi con valori al di fuori del minimo e del massimo specificato saranno esclusi dal test. Se, ad esempio, si specifica 1 per il limite inferiore e 4 per il limite superiore, per il test del chi-quadrato verranno utilizzati solo i valori interi compresi tra 1 e 4.

Valori previsti. Per impostazione predefinita, tutte le categorie hanno valori attesi uguali. Per le categorie sono previste proporzioni attese definite dall'utente. Selezionare **Valori**, specificare un valore maggiore di 0 per ogni categoria della variabile del test e quindi fare clic su **Aggiungi**. I valori vengono elencati in ordine di inserimento, L'ordine dei valori è importante in quanto corrisponde all'ordine crescente dei valori delle categorie della variabile oggetto del test. Il primo valore dell'elenco corrisponde al valore di gruppo minore della variabile, mentre l'ultimo corrisponde al valore maggiore. Gli elementi dell'elenco dei valori vengono sommati e quindi ciascun valore viene diviso per la somma risultante per calcolare la proporzione di casi attesi nella categoria corrispondente. Ad esempio, un elenco valori formato da 3, 4, 5, 4 specifica le proporzioni attese 3/16, 4/16, 5/16 e 4/16.

Test del chi-quadrato: Opzioni

Statistiche. È possibile scegliere una o entrambe le statistiche di riepilogo.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Escludi casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Escludi casi listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (test del chi-quadrato)

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare valori minimi e massimi diversi o frequenze attese diverse per variabili diverse (con il sottocomando CHISQUARE).
- Eseguire il test confrontando la stessa variabile con diverse frequenze attese o utilizzando diversi intervalli (con il sottocomando EXPECTED).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test binomiale

Grazie alla procedura del test binomiale è possibile confrontare le frequenze osservate delle due categorie di una variabile dicotomica con le frequenze previste in presenza di una distribuzione binomiale con il parametro di probabilità specificato. Per impostazione predefinita, il parametro di probabilità per entrambi i gruppi è 0,5. Per modificare le probabilità, è possibile inserire una proporzione di test per il primo gruppo. La probabilità per il secondo gruppo sarà uguale a 1 meno la probabilità specificata per il primo gruppo.

Esempio. Quando si lancia una monetina, la probabilità che esca testa è uguale a 1/2. In base a questa ipotesi, una monetina viene lanciata 40 volte e i risultati vengono registrati (testa o croce). Dal test binomiale può risultare che per i 3/4 dei lanci della moneta è uscita testa e che il livello di significatività è molto basso (0,0027). Questi risultati indicano che la probabilità che venga testa molto spesso non è pari a 1/2; pertanto, la stima sul comportamento della moneta probabilmente risulta distorta.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili.

Considerazioni sui dati di test binomiale

Dati. Le variabili incluse nel test devono essere numeriche e dicotomiche. Per convertire le variabili stringa in variabili numeriche, utilizzare il comando Ricodifica automatica del menu Trasforma. Una **variabile dicotomica** è una variabile che può avere solo due valori possibili: *sì* o *no*, *true* o *false*, 0 o 1, e così via. Il primo valore rilevato nel dataset definisce il primo gruppo, mentre l'altro valore definisce il secondo gruppo. Se le variabili sono dicotomiche, è necessario specificare un punto di divisione. Utilizzando il punto di divisione è possibile assegnare al primo gruppo i casi con valori inferiori o uguali al punto di divisione e i rimanenti casi al secondo gruppo.

Ipotesi. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Si presume che i dati rappresentino un campione casuale.

Per ottenere un test binomiale

1. Dai menu, scegliere:

   Analizza > **Test non parametrici** > **Finestre di dialogo legacy** > **Binomiale...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili numeriche oggetto del test.
3. È possibile fare clic su **Opzioni** per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

Test binomiale: Opzioni

Statistiche. È possibile scegliere una o entrambe le statistiche di riepilogo.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Escludi casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Escludi casi listwise.** I casi con valori mancanti per qualsiasi variabile da verificare verranno esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TEST (test binomiale)

Il linguaggio della sintassi dei comandi consente inoltre di:

- Selezionare gruppi specifici (ed escluderne altri) quando una variabile ha più di due categorie (con il sottocomando `BINOMIAL`).
- Specificare diversi punti di divisione o probabilità per variabili diverse (con il sottocomando `BINOMIAL`).
- Eseguire test confrontando la stessa variabile con diversi punti di divisione o probabilità (con il sottocomando `EXPECTED`).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test delle successioni

Il test delle successioni verifica se l'ordine delle occorrenze di due valori di una variabile è casuale. Una successione è una sequenza di osservazioni simili. Un campione con troppe o troppo poche successioni indica che il campione non è casuale.

Esempi. Si supponga che a venti persone venga chiesto se comprerebbero un determinato prodotto. La casualità prevista per il campione viene messa fortemente in dubbio se tutte le venti persone sono dello stesso sesso. È possibile utilizzare il test delle successioni per determinare se il campione è stato estratto in modo casuale.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili.

Considerazioni sui dati di test delle successioni

Dati. Le variabili devono essere numeriche. Per convertire le variabili stringa in variabili numeriche, utilizzare il comando Ricodifica automatica del menu Trasforma.

Ipotesi. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Utilizzare campioni da distribuzioni di probabilità continue.

Per ottenere un test delle successioni

1. Dai menu, scegliere:

   Analizza > **Test non parametrici** > **Finestre di dialogo legacy** > **Esecuzioni...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili numeriche oggetto del test.
3. È possibile fare clic su **Opzioni** per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

Test delle successioni: Punto di divisione

Punto di divisione. Specifica un punto di divisione per dicotomizzare le variabili scelte dall'utente. È possibile utilizzare la media osservata, la mediana o la moda oppure un valore specificato come punto di divisione. I casi con valori minori del punto di divisione sono assegnati a un gruppo e i casi con valori uguali o maggiori del punto di divisione sono assegnati a un altro gruppo. Viene eseguito un test per ogni punto di divisione selezionato.

Opzioni del test delle successioni

Statistiche. È possibile scegliere una o entrambe le statistiche di riepilogo.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Escludi casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Escludi casi listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (Test delle successioni)

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare diversi punti di divisione per diverse variabili (con il sottocomando RUNS).
- Verificare la stessa variabile rispetto a diversi punti di divisione personalizzati (con il sottocomando RUNS).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test di Kolmogorov-Smirnov a campione singolo

La procedura Test di Kolmogorov-Smirnov per un campione consente di confrontare la funzione di distribuzione cumulativa osservata per una variabile con la distribuzione teorica specificata, che può essere normale, uniforme o di Poisson. La *Z* di Kolmogorov-Smirnov viene calcolata in base alla differenza maggiore (in valore assoluto) tra la funzione di distribuzione cumulativa osservata e teorica. Questo test sulla bontà di adattamento permette di verificare se le osservazioni possono provenire dalla distribuzione specificata.

A partire dalla versione 27.0, la statistica del test di Lilliefors può essere utilizzata per stimare il valore *p*utilizzando il campionamento di Monte Carlo per il test rispetto a una distribuzione normale con parametri stimati (questa funzionalità era precedentemente possibile solo tramite la procedura Esplora).

Esempio
Molti test parametrici richiedono variabili distribuite in modo normale. Il test di Kolmogorov-Smirnov per un campione può essere utilizzato per verificare che una variabile, ad esempio *reddito*, sia distribuita in modo normale.

Statistiche
Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti, quartili, test di Lilliefors e simulazione Monte Carlo.

Considerazioni sui dati di test di Kolmogorov-Smirnov per un campione

Dati
Utilizzare variabili quantitative (misurazione a livello di intervallo o di rapporto).

Ipotesi
Il test di Kolmogorov-Smirnov presume che i parametri della distribuzione del test vengano specificati anticipatamente. Questa procedura consente di valutare i parametri del campione. La media e la deviazione standard del campione sono i parametri della distribuzione normale, i valori minimo e massimo del campione definiscono l'intervallo di distribuzione uniforme e la media del campione è il parametro per la distribuzione Poisson e per la distribuzione esponenziale. La capacità del test di rilevare gli scostamenti dalla distribuzione ipotizzata può essere seriamente compromessa.

Quando alcuni parametri della distribuzione devono essere stimati dal campione, il test Kolmogorov-Smirnov non si applica più. In queste istanze, la statistica di prova di Lilliefors può essere utilizzata per stimare il valore *p*utilizzando il campionamento di Monte Carlo per la verifica della normalità con media e varianza sconosciuta. Il test di Lilliefors si applica alle tre distribuzioni continue (**Normal**, **Exponentia**l**e Uniform**). Si noti che il test non si applica se la distribuzione sottostante è discreta (**Poisson**). Il test è definito solo per l'inferenza di un campione quando i parametri di distribuzione corrispondenti non sono specificati.

Ottenimento di un test di One - Sample Kolmogorov-Smirnov

1. Dai menu, scegliere:

   Analizza > **Test non parametrici** > **Finestre di dialogo legacy** > **K-S per 1 campione...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili numeriche oggetto del test. Ogni variabile produce un test distinto.
3. Opzionalmente, selezionare un metodo di distribuzione del test:

   Normale
   Quando viene selezionato, è possibile specificare se i parametri di distribuzione sono stimati dai dati di esempio (l'impostazione predefinita) o da impostazioni personalizzate. Quando si seleziona **Utilizza dati di esempio** , vengono utilizzati sia i risultati asintotici esistenti che la correzione di significatività di Lilliefors basata sul campionamento di Monte Carlo. Quando **Custom** viene selezionato, fornire valori sia per **Mean** che **Std Dev**.

Uniforme
Quando viene selezionato, è possibile specificare se i parametri di distribuzione sono stimati dai dati di esempio (l'impostazione predefinita) o da impostazioni personalizzate. Quando **Utilizza dati di esempio** viene selezionato, viene utilizzato il test di Lilliefors. Quando **Custom** viene selezionato, fornire valori sia per **Mina** che **Max**.

Poisson
Quando selezionato, specificare un valore di parametro **Media**.

Esponenziale
Quando viene selezionato, è possibile specificare se i parametri di distribuzione sono stimati dalla media del campione (l'impostazione predefinita) o da impostazioni personalizzate. Quando **Utilizza dati di esempio** viene selezionato, viene utilizzato il test di Lilliefors. Quando **Custom** viene selezionato, fornire un valore di parametro **Mean** .

4. Opzionalmente, clicca su **Simulazione** per specificare i parametri di simulazione di Monte Carlo, fare clic su **Exact** per specificare i parametri di prova esatti o fare clic su **Opzioni** per le statistiche descrittive, le quartili e il controllo del trattamento dei dati mancanti.

One - Esempio Kolmogorov-Smirnov Test: Simulazione

Quando alcuni parametri della distribuzione devono essere stimati dal campione, il test Kolmogorov-Smirnov non si applica più. In queste istanze, la statistica di prova di Lilliefors può essere utilizzata per stimare il valore *p*utilizzando il campionamento di Monte Carlo per la verifica della normalità con media e varianza sconosciuta. Il test di Lilliefors si applica alle tre distribuzioni continue (**Normal**, **Exponentia**le **Uniform**). Si noti che il test non si applica se la distribuzione sottostante è discreta (**Poisson**). Il test è definito solo per l'inferenza di un campione quando i parametri di distribuzione corrispondenti non sono specificati.

Parametri della simulazione Monte Carlo

Livello di confidenza
Questa impostazione opzionale reimposta il livello di intervallo di confidenza stimato dal test Kolmogorov-Smirnov quando si utilizzano le simulazioni Monte Carlo. Il valore deve essere compreso tra 0 e 100. L'impostazione predefinita è 99.

Numero di campioni
Questa impostazione opzionale reimposta il numero di repliche che il test di Lilliefors utilizza per il campionamento di Monte Carlo. Il valore deve essere un intero intero compreso tra 10000 e il maggior numero di campioni di valore. Il valore predefinito è 10000.

Sopprimi i risultati Monte Carlo per la distribuzione normale
Questa impostazione opzionale sopprime il campionamento di Monte Carlo per i normali risultati di distribuzione. Per impostazione predefinita, l'impostazione non viene selezionata (il che significa sia i risultati asintotici esistenti che i risultati del test di Lilliefors, che si basano sul campionamento di Monte Carlo, sono presentati).

Test di Kolmogorov-Smirnov a campione singolo: Opzioni

Statistiche
È possibile scegliere una o entrambe le statistiche di riepilogo.

Descrittive
Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.

Quartili
Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti
Consente di controllare la modalità di elaborazione dei valori mancanti.

Escludi casi test per test
Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.

Escludi casi listwise
I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (Test di Kolmogorov-Smirnov per un campione)

Il linguaggio della sintassi dei comandi consente anche di specificare i parametri della distribuzione del test (con il sottocomando K-S).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test di due campioni indipendenti

La procedura del test per due campioni indipendenti consente di confrontare due gruppi di casi in base a una sola variabile.

Esempio. Sono stati creati nuovi apparecchi odontoiatrici che presentano numerosi vantaggi in termini di comodità, estetica ed efficacia ai fini dell'allineamento dei denti. Per determinare se i nuovi e i vecchi apparecchi devono essere portati per lo stesso periodo, sono stati scelti casualmente 10 bambini con il vecchio apparecchio e 10 bambini con il nuovo apparecchio. Dal test *U* di Mann-Whitney è possibile riscontrare che in media il nuovo apparecchio deve essere portato per un periodo di tempo inferiore.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: *U* di Mann-Whitney, Reazioni estreme di Moses, *Z* di Kolmogorov-Smirnov, successioni di Wald-Wolfowitz.

Considerazioni sui dati per i test per due campioni indipendenti

Dati. Utilizzare variabili numeriche che possono essere ordinate.

Ipotesi. Utilizzare campioni casuali indipendenti. Il test *U* di Mann-Whitney verifica l'uguaglianza di due distribuzioni. Per poterlo utilizzare per verificare le differenze nell'ubicazione di due distribuzioni, è necessario presupporre che le distribuzioni abbiano la stessa forma.

Per ottenere un test per due campioni indipendenti

1. Dai menu, scegliere:

   Analizza > **Test non parametrici** > **Finestre di dialogo legacy** > **2 campioni indipendenti...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili numeriche.
3. Selezionare una variabile di raggruppamento e fare clic su **Definisci gruppi** per suddividere il file in due gruppi o campioni.

Tipi di test per due campioni indipendenti

Tipo di test. Sono disponibili quattro test che consentono di verificare se due campioni (gruppi) indipendenti provengono dalla stessa popolazione.

Il **test U di Mann-Whitney** è il test per due campioni indipendenti più diffuso. Equivale al test della somma dei ranghi di Wilcoxon e al test di Kruskal-Wallis per due gruppi. Il test di Mann-Whitney permette di verificare l'equivalenza dell'ubicazione delle due popolazioni campione. Le osservazioni di entrambi i gruppi vengono combinate e ordinate per ranghi, assegnando il rango medio in caso di correlazioni. Il numero di correlazioni deve essere inferiore al numero totale di osservazioni. Se l'ubicazione delle popolazioni risulta identica, è necessario distribuire casualmente i ranghi tra i due campioni. Il test calcola il numero di volte che un punteggio del gruppo 1 precede un punteggio del gruppo 2 e il numero di volte che un punteggio del gruppo 2 precede un punteggio del gruppo 1. La statistica Mann - Whitney *U* è la più piccola di questi due numeri. Viene visualizzata anche la statistica *W* della somma dei ranghi di Wilcoxon. *W* è la somma dei ranghi per il gruppo con la classifica media più piccola, a meno che i gruppi non abbiano lo stesso rango medio, nel qual caso si tratta della somma di rango del gruppo che si chiama ultima nella finestra di dialogo Two - Independent - Samples Definisci Groups.

Il **test Z di Kolmogorov-Smirnov** e il **test delle successioni di Wald-Wolfowitz** sono test di carattere più generale che consentono di individuare le differenze tra le distribuzioni in termini di forma e ubicazione. Il test di Kolmogorov-Smirnov si basa sulla massima differenza in valore assoluto tra le funzioni di distribuzione cumulative osservate per entrambi i campioni. Quando tale differenza è significativa, le due distribuzioni vengono considerate diverse. Il test delle successioni di Wald-Wolfowitz consente di combinare e ordinare in ranghi le osservazioni di entrambi i gruppi. Se i due campioni provengono dalla stessa popolazione, è necessario distribuire casualmente i gruppi all'interno della classificazione.

Il **test delle reazioni estreme di Moses** si basa sull'ipotesi che la variabile sperimentale influenzi alcuni soggetti in una direzione e altri nella direzione opposta. Consente di verificare le risposte estreme confrontandole con un gruppo di controllo. Questo test è incentrato sull'estensione del gruppo di controllo e definisce la misura in cui i valori estremi del gruppo sperimentale influenzano l'estensione in caso di combinazione con il gruppo di controllo. Il gruppo di controllo viene definito dal valore del gruppo 1 nella finestra di dialogo Due campioni indipendenti: Definisci gruppi. Le osservazioni eseguite su entrambi i gruppi vengono combinate e classificate per ranghi. L'arco del gruppo di controllo viene calcolato come differenza tra i ranghi dei valori più grandi e più piccoli nel gruppo di controllo più 1. Poiché gli outliers di possibilità possono snaturare facilmente la gamma dell'arco, il 5% dei casi di controllo viene rifilato automaticamente da ogni estremità.

Test per due campioni indipendenti: Definisci gruppi

Per dividere il file in due gruppi o campioni, immettere un valore intero per Gruppo 1 e un altro valore per Gruppo 2. I casi con altri valori vengono esclusi dall'analisi.

Test per due campioni indipendenti: Opzioni

Statistiche. È possibile scegliere una o entrambe le statistiche di riepilogo.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Escludi casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Escludi casi listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (Due campioni indipendenti)

Il linguaggio della sintassi dei comandi permette anche di specificare il numero di casi da ritagliare dal test di Moses (con il sottocomando MOSES).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test per due campioni correlati

La procedura del test per due campioni dipendenti consente di confrontare la distribuzione di due variabili.

Esempio. In generale, le famiglie ricevono l'intero prezzo di offerta per la vendita della propria casa? Applicando il test dei ranghi con segno di Wilcoxon ai dati relativi a 10 case, si riscontrerà che sette famiglie ricevono una somma inferiore al prezzo di offerta, una famiglia riceve una somma superiore, mentre due sole famiglie lo ricevono interamente.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: ranghi con segni di Wilcoxon, del segno, McNemar. Se è installato il modulo Exact Tests (disponibile solo nei sistemi operativi Windows), è disponibile anche il test di omogeneità marginale.

Considerazioni sui dati per un test per due campioni dipendenti

Dati. Utilizzare variabili numeriche che possono essere ordinate.

Ipotesi. Anche se per le due variabili non si ipotizza una particolare distribuzione, si presume che la distribuzione delle differenze a coppie sia simmetrica.

Per ottenere test per due campioni dipendenti

1. Dai menu, scegliere:

   Analizza > **Test non parametrici** > **Finestre di dialogo legacy** > **2 campioni correlati...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più coppie di variabili.

Tipi di test per due campioni dipendenti

I test descritti in questa sezione permettono di confrontare le distribuzioni di due variabili correlate. Il test più appropriato varia a seconda dei tipi di dati.

Se i dati sono continui, utilizzare il test dei segni o dei ranghi con segno di Wilcoxon. Il **test dei segni** permette di calcolare le differenze tra le due variabili per tutti i casi e di classificarle come positive, negative o correlate. Se le due variabili sono distribuite in modo analogo, il numero di differenze positive e negative non differirà in misura significativa. Il **test dei ranghi con segno di Wilcoxon** prende in considerazione le informazioni relative al segno e alla grandezza delle differenze tra le coppie. Poiché il test dei ranghi con segno di Wilcoxon include un maggior numero di informazioni relative ai dati, risulta più valido del test dei segni.

Se i dati sono binari, utilizzare il **test di McNemar**. Questo test viene in genere utilizzato in presenza di misure ripetute, ovvero quando la risposta del soggetto viene richiesta due volte: prima e dopo il verificarsi di un determinato evento. Il test di McNemar consente di determinare se il tasso di risposta iniziale (prima dell'evento) equivale al tasso di risposta finale (dopo l'evento). Questo test risulta particolarmente utile per individuare le variazioni della risposta in disegni sperimentali del tipo 'prima e dopo'.

Se i dati sono categoriali, utilizzare il **test di omogeneità marginale**. Estensione del test di McNemar dalla risposta binaria a quella multinomiale. Consente di verificare le variazioni della risposta utilizzando la distribuzione del chi-quadrato e risulta utile in disegni sperimentali del tipo 'prima e dopo'. Il test di omogeneità marginale è disponibile solo se è stato installato il modulo Exact Tests.

Test per due campioni dipendenti: Opzioni

Statistiche. È possibile scegliere una o entrambe le statistiche di riepilogo.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Escludi casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Escludi casi listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (due campioni dipendenti)

Il linguaggio della sintassi dei comandi permette anche di verificare una variabile con ciascuna variabile dell'elenco.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test per diversi campioni indipendenti

La procedura per i test per diversi campioni indipendenti consente di confrontare due o più gruppi di casi in base a una variabile.

Esempio. Le lampadine da 100 watt di tre diversi produttori si differenziano in relazione al tempo medio di bruciatura del filamento? Grazie all'analisi della varianza a una via di Kruskal-Wallis è possibile verificare che la durata media delle tre lampadine è effettivamente diversa.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: *H* di Kruskal-Wallis, mediana.

Considerazioni sui dati dei test per diversi campioni indipendenti

Dati. Utilizzare variabili numeriche che possono essere ordinate.

Ipotesi. Utilizzare campioni casuali indipendenti. IL test *H* di Kruskal-Wallis richiede che i campioni sottoposti a test siano simili per forma.

Per ottenere test per diversi campioni indipendenti

1. Dai menu, scegliere:

   Analizza > **Test non parametrici** > **Finestre di dialogo legacy** > **K campioni indipendenti...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili numeriche.
3. Selezionare una variabile di raggruppamento e fare clic su **Definisci intervallo** per specificare i valori interi minimo e massimo per la variabile di raggruppamento.

Test per diversi campioni indipendenti: tipi di test

Sono disponibili tre test per stabilire se diversi campioni indipendenti sono stati estratti dalla stessa popolazione. Il test *H* di Kruskal-Wallis, il test della mediana e il test di Jonckheere-Terpstra consentono di verificare se i diversi campioni indipendenti sono stati estratti dalla stessa popolazione.

Il test **H di Kruskal-Wallis**, un'estensione del test *U* di Mann-Whitney, è la versione non parametrica dell'analisi della varianza a una via e consente di rilevare le differenze nell'ubicazione di distribuzione. Il **test della mediana**, che è più generale ma non altrettanto potente, consente di rilevare le differenze distribuzionali nell'ubicazione e nella forma. Il test *H* di Kruskal-Wallis e il test della mediana presumono che non esistano ordinamenti *a priori* delle *k* popolazioni da cui sono estratti i campioni.

Quando *esiste* un naturale ordinamento *a priori* (crescente o decrescente) delle *k* popolazioni, il **test di Jonckheere-Terpstra** è più potente. Ad esempio, le *k* popolazioni possono rappresentare *k* temperature crescenti. L'ipotesi che diverse temperature producano la stessa distribuzione della risposta è verificata rispetto all'ipotesi alternativa in base a cui al salire della temperatura, cresce la grandezza della risposta. Qui l'ipotesi alternativa è ordinata e quindi il test di Jonckheere-Terpstra è il più appropriato da utilizzare. Il test di Jonckheere-Terpstra è disponibile solo se è installato il modulo aggiuntivo Exact Tests.

Test per diversi campioni indipendenti: Definisci intervallo

Per definire l'intervallo, immettere valori interi per il **minimo** e il **massimo** che corrispondono alle categorie minore e maggiore della variabile di raggruppamento. Sono esclusi i casi con valori al di fuori dei limiti. Se, ad esempio, si specifica un limite inferiore di 1 e un limite superiore di 3, verranno utilizzati solo i valori interi compresi tra 1 e 3. Il valore minimo deve essere inferiore al valore massimo ed entrambi i valori devono essere specificati.

Test per diversi campioni indipendenti: Opzioni

Statistiche. È possibile scegliere una o entrambe le statistiche di riepilogo.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Escludi casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Escludi casi listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (K campioni indipendenti)

Il linguaggio della sintassi dei comandi permette anche di specificare un valore diverso dalla mediana osservata per il test della mediana (con il sottocomando `MEDIAN`).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Test per diversi campioni correlati

Il test per diversi campioni dipendenti consente di confrontare le distribuzioni di due o più variabili.

Esempio. Il pubblico associa diversi livello di prestigio al ruolo di dottore, avvocato, ufficiale della polizia e insegnante? A dieci persone viene chiesto di classificare queste quattro occupazioni in base al prestigio. Il test di Friedman indica che il pubblico associa effettivamente livelli di prestigio diversi a queste quattro professioni.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: di Friedman, *W* di Kendall e *Q* di Cochran.

Considerazioni sui dati relativi ai test per diversi campioni dipendenti

Dati. Utilizzare variabili numeriche che possono essere ordinate.

Ipotesi. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Utilizzare campioni casuali dipendenti.

Per ottenere i test per diversi campioni dipendenti

1. Dai menu, scegliere:

   Analizza > **Test non parametrici** > **Finestre di dialogo legacy** > **K campioni correlati...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili oggetto del test numeriche.

Test per diversi campioni dipendenti: tipi di test

Sono disponibili tre test per confrontare le distribuzioni di diverse variabili correlate.

Il **test di Friedman** è l'equivalente non parametrico di un disegno di misure ripetute per un campione o ANOVA a due vie con una osservazione per cella. Friedman verifica l'ipotesi nulla secondo cui *k* variabili correlate provengono dalla stessa popolazione. Per ogni caso, le variabili *k* vengono ordinate da 1 a *k*. La statistica di prova si basa su questi ranghi.

W di Kendall è una normalizzazione delle statistiche di Friedman. È possibile interpretare il *W* di Kendall come il coefficiente di concordanza, che rappresenta la misura dell'accordo tra stimatori. Ogni caso è uno stimatore e ogni variabile è un elemento o individuo da stimare. Per ogni variabile viene calcolata la somma dei ranghi. *W* di Kendall varia tra 0 (nessun accordo) e 1 (accordo completo).

Q di Cochran è identico al test di Friedman ma è applicabile quando tutte le risposte sono binarie. Questo test è un'estensione del test di McNemar alla situazione di *k*-campioni. I test *Q* di Cochran verificano l'ipotesi secondo cui diverse variabili dicotomiche hanno la stessa media. Le variabili sono misurate sullo stesso individuo o su individui collegati fra loro.

Test per diversi campioni dipendenti: Statistica

È possibile scegliere le statistiche.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.

- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Funzioni aggiuntive del comando NPAR TESTS (K campioni dipendenti)

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Analisi a risposta multipla

Analisi a risposta multipla

Sono disponibili due procedure per l'analisi di insiemi a dicotomie e a categorie multiple. La procedura Risposte multiple: Frequenze consente di visualizzare le tabelle delle frequenze. La procedura Risposte multiple: Tabelle di contingenza consente di visualizzare tavole di contingenza a due o a tre dimensioni. Prima di utilizzare una delle procedure descritte, è necessario definire gli insiemi a risposta multipla.

Esempio. In questo esempio viene illustrato l'utilizzo degli elementi a risposta multipla in un'indagine di mercato. I dati sono fittizi e non devono essere interpretati come reali. È possibile condurre un'indagine tra i passeggeri di una linea aerea in volo su una particolare rotta per ottenere una valutazione della concorrenza. In questo esempio la compagnia American Airlines conduce un'indagine volta a rilevare se i propri passeggeri viaggiano con altre linee aeree sulla rotta Chicago-New York e a determinare l'importanza relativa dei fattori di programmazione e di servizio ai fini della scelta della linea aerea. Al momento dell'imbarco, l'assistente di volo consegna a ciascun passeggero un breve questionario. Nella prima domanda si legge: cerchiare tutte le compagnie aeree, tra le seguenti, con le quali si è volato almeno una volta negli ultimi sei mesi in questa rotta--American, United, TWA, USAir, altro. Si tratta di una domanda a risposta multipla in quanto il passeggero può indicare più risposte. La domanda, tuttavia, non può essere codificata direttamente in quanto una variabile può contenere un solo valore per ciascun caso. È necessario utilizzare più variabili per associare le risposte a ciascuna domanda. Per eseguire questa operazione è possibile procedere in due modi. Il primo modo consiste nel definire una variabile per ciascuna delle scelte (ad esempio, American, United, TWA, USAir e Altro). Se gli ambienti passeggeri Uniti, la variabile *unita* viene assegnato un codice di 1, altrimenti 0. Questo è un **metodo dichotomy multiplo** di variabili di mappatura. Il secondo metodo per la classificazione delle risposte è il **metodo a categorie multiple**, che consente di valutare il numero massimo di risposte possibili alla domanda e di impostare lo stesso numero di variabili, con i codici utilizzati per specificare la linea aerea utilizzata. Dall'esame di un campione di questionari può risultare che negli ultimi sei mesi nessun utente ha viaggiato con più di tre diverse linee aeree su questa rotta. Può inoltre risultare che, a causa della deregulation delle linee aree, nella categoria Altro ne vengano citate altre 10. Utilizzando il metodo a risposta multipla, vengono definite tre variabili, ciascuna delle quali è codificata come 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta* e così via. Se un determinato passeggero indica American e TWA, alla prima variabile viene assegnato il codice 1, alla seconda il codice 3 e alla terza un codice di valore mancante. Un altro passeggero può aver indicato American e specificato Delta. In questo caso, alla prima variabile viene assegnato il codice 1, alla seconda 5 e alla terza un codice di valore mancante. Se invece si utilizza il metodo a dicotomie multiple, si otterranno 14 variabili distinte. Sebbene ai fini di questa indagine sia possibile utilizzare entrambi i metodi di mapping, la scelta del metodo dipende dalla distribuzione delle risposte.

Risposte multiple: Definisci insiemi

La procedura di definizione degli insiemi di variabili a risposta multipla consente di raggruppare le variabili elementari in insiemi a dicotomie o a categorie multiple, per i quali è possibile ottenere tabelle delle frequenze e tavole di contingenza. È possibile definire fino a 20 insiemi a risposta multipla. A ogni insieme è necessario assegnare un nome univoco. Per rimuovere un insieme, evidenziarlo nell'elenco dei gruppi a risposta multipla e quindi scegliere **Rimuovi**. Per modificare un insieme, evidenziarlo nell'elenco, modificare le caratteristiche di definizione desiderate e quindi scegliere **Cambia**.

È possibile codificare le variabili elementari come dicotomie o categorie. Per l'utilizzo delle variabili dicotomiche, selezionare **Dicotomie** per creare un insieme a dicotomie multiple. Specificare un valore intero nella casella Valore conteggiato. Ciascuna variabile contenente almeno un'occorrenza del valore conteggiato diventa una categoria dell'insieme a dicotomie multiple. Selezionare **Categorie** per creare un insieme a categorie multiple con lo stesso intervallo di valori delle variabili che lo compongono.

Specificare valori interi come valori minimo e massimo dell'intervallo di categorie dell'insieme a categorie multiple. Verrà calcolato il totale di ogni singolo valore intero nell'intervallo per tutte le variabili. Le categorie vuote non verranno ordinate in tabelle.

A ogni insieme a risposta multipla deve essere assegnato un nome univoco composto al massimo da sette caratteri. Al nome assegnato verrà aggiunto automaticamente il prefisso $ (segno di dollaro). Non è possibile utilizzare i seguenti nomi riservati: *casenum*, *sysmis*, *jdate*, *date*, *time*, *length* e *width*. Il nome dell'insieme a risposta multipla esiste solo ai fini dell'utilizzo in procedure a risposta multipla. Non è possibile fare riferimento ai nomi di insiemi a risposta multipla in altre procedure. È inoltre possibile inserire un'etichetta descrittiva di variabile per l'insieme a risposta multipla. L'etichetta può essere costituita al massimo da 40 caratteri.

Per definire gli insiemi a risposta multipla

1. Dai menu, scegliere:

   Analizza > **Risposte multiple** > **Definisci insiemi di variabili...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare due o più variabili.
3. Se le variabili sono codificate come dicotomie, indicare il valore che si desidera calcolare. Se le variabili sono codificate come categorie, definire l'intervallo delle categorie.
4. Immettere un nome univoco per ciascun insieme a risposta multipla.
5. Scegliere **Aggiungi** per aggiungere l'insieme a risposta multipla all'elenco di insiemi definiti.

Risposte multiple: Frequenze

La procedura Risposte multiple: Frequenze consente di ottenere tabelle delle frequenze per gli insiemi a risposta multipla. È innanzitutto necessario definire uno o più insiemi a risposta multipla (vedere "Risposte multiple: Definisci insiemi").

Per gli insiemi a dicotomie multiple, i nomi delle categorie indicati nell'output vengono determinati in base alle etichette definite per le variabili elementari del gruppo. Se le etichette di variabile non sono definite, i nomi delle variabili verranno utilizzati come etichette. Per gli insiemi a categorie multiple, le etichette di categoria vengono determinate in base alle etichette valore della prima variabile del gruppo. Se le categorie mancanti per la prima variabile sono presenti per altre variabili del gruppo, definire un'etichetta valore per le categorie mancanti.

Valori mancanti. I casi con valori mancanti vengono esclusi tabella per tabella. È inoltre possibile scegliere una delle seguenti opzioni o entrambe:

- **Escludi casi listwise all'interno delle dicotomie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabulazione dell'insieme a dicotomie multiple. Questa opzione può essere applicata solo agli insiemi a risposta multipla definiti come insiemi dicotomici. Per impostazione predefinita, un caso viene considerato mancante per un insieme a dicotomie multiple se nessuna delle variabili che lo compongono contiene il valore conteggiato. I casi con valori mancanti solo per alcune variabili verranno inclusi nelle tabulazioni del gruppo se almeno una variabile contiene il valore conteggiato.
- **Escludi casi listwise all'interno delle categorie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabulazione dell'insieme a categorie multiple. Questa opzione viene applicata solo a insiemi a risposta multipla definiti come insiemi di categorie. Per impostazione predefinita, un caso viene considerato mancante per un insieme a categorie multiple solo se nessuno dei componenti contiene valori validi all'interno dell'intervallo definito.

Esempio. Ogni variabile creata in base a una domanda di un'indagine è una variabile elementare. Per analizzare un elemento a risposta multipla, è necessario combinare le variabili in uno dei due tipi di insiemi a risposta multipla; un insieme a dicotomie multiple o un insieme a categorie multiple. Se ad esempio in un'indagine sulle linee aeree dove viene richiesto con quale delle tre linee aeree indicate (American, United, TWA) si è viaggiato negli ultimi sei mesi sono state utilizzate variabili dicotomiche

e si è definito un **insieme a dicotomie multiple**, ciascuna delle tre variabili dell'insieme diventerà una categoria della variabile di gruppo. I conteggi e le percentuali relativi alle tre linee aeree verranno visualizzati in una sola tabella delle frequenze. Se risulta che nessun rispondente ha indicato più di due linee aeree, è possibile creare due variabili, ciascuna con tre codici, ovvero uno per ogni linea aerea. Se si definisce un **insieme a categorie multiple**, i valori verranno ordinati in tabelle aggiungendo gli stessi codici alle variabili elementari. L'insieme di valori risultante equivale agli insiemi di ciascuna variabile elementare. Ad esempio, 30 risposte per United sono la somma delle cinque risposte United per la compagnia aerea 1 e le 25 risposte United per la compagnia aerea 2. I conteggi e le percentuali relativi alle tre linee aeree verranno visualizzati in una sola tabella delle frequenze.

Statistiche. Tabelle delle frequenze in cui vengono visualizzati i conteggi, le percentuali delle risposte, le percentuali dei casi, il numero di casi validi e il numero di casi mancanti.

Considerazioni sui dati relativi alle frequenze delle risposte multiple

Dati. Utilizzare gli insiemi a risposta multipla.

Ipotesi. I conteggi e le percentuali forniscono un'utile descrizione dei dati provenienti da qualsiasi distribuzione.

Procedure correlate. La procedura Risposte multiple: Definisci insiemi consente di definire insiemi a risposta multipla.

Per ottenere le frequenze delle risposte multiple

1. Dai menu, scegliere:

   Analizza > **Risposte multiple** > **Frequenze...**
2. Selezionare uno o più insiemi a risposta multipla.

Risposte multiple: Tabelle di contingenza

La procedura Risposte multiple: Tabelle di contingenza consente di incrociare insiemi a risposta multipla definiti, variabili elementari o una combinazione di entrambi. È inoltre possibile ottenere le percentuali delle celle in base a casi o risposte, modificare il trattamento dei valori mancanti oppure ottenere tavole di contingenza accoppiate. È innanzitutto necessario definire uno o più insiemi a risposta multipla (vedere "Per definire gli insiemi a risposta multipla").

Per gli insiemi a dicotomie multiple, i nomi delle categorie indicati nell'output vengono determinati in base alle etichette definite per le variabili elementari del gruppo. Se le etichette di variabile non sono definite, i nomi delle variabili verranno utilizzati come etichette. Per gli insiemi a categorie multiple, le etichette di categoria vengono determinate in base alle etichette valore della prima variabile del gruppo. Se le categorie mancanti per la prima variabile sono presenti per altre variabili del gruppo, definire un'etichetta valore per le categorie mancanti. Le etichette delle categorie per le colonne verranno visualizzate su tre righe, composte al massimo da otto caratteri ciascuna. Per evitare di suddividere le parole, è possibile invertire le righe e le colonne oppure ridefinire le etichette.

Esempio. In questa procedura è possibile incrociare gli insiemi a dicotomie e a categorie multiple con altre variabili. In un sondaggio destinato ai passeggeri di una linea aerea vengono chieste ai passeggeri le seguenti informazioni: cerchiare tutte le compagnie aeree, tra le seguenti, con le quali si è volato almeno una volta negli ultimi sei mesi (American, United, TWA). Cosa è più importante per la scelta di un volo, la programmazione o il servizio offerto? Scegliere una sola opzione. Dopo aver inserito i dati come dicotomie o categorie multiple e averli uniti in un insieme, è possibile incrociare le domande relative alle linee aeree con quelle relative al servizio o alla programmazione.

Statistiche. Tavola di contingenza con conteggi relativi a celle, righe, colonne e totale e percentuali relative a celle, righe, colonne e totale. Le percentuali delle celle possono basarsi sui casi o sulle risposte.

Considerazioni sui dati relativi alle tabelle di contingenza a risposta multipla

Dati. Utilizzare insiemi a risposta multipla oppure variabili categoriali numeriche.

Ipotesi. I conteggi e le percentuali forniscono un'utile descrizione dei dati provenienti da qualsiasi distribuzione.

Procedure correlate. La procedura Risposte multiple: Definisci insiemi consente di definire insiemi a risposta multipla.

Per ottenere tabelle di contingenza a risposta multipla

1. Dai menu, scegliere:

   Analizza > **Risposte multiple** > **Tabelle di contingenza...**
2. Selezionare una o più variabili numeriche o insiemi a risposta multipla per ciascuna dimensione della tavola di contingenza.
3. Definire l'intervallo di ciascuna variabile elementare.

È inoltre possibile ottenere una tavola di contingenza a due vie per ciascuna categoria di una variabile di controllo o di un insieme a risposta multipla. Selezionare uno o più elementi dall'elenco Strati.

Risposte multiple, tabelle di contingenza: Definisci intervalli delle variabili

È necessario definire gli intervalli dei valori per tutte le variabili elementari nella tavola di contingenza. Specificare il valore di categoria minimo e massimo intero che si desidera ordinare in tabelle. Le categorie che non rientrano nell'intervallo verranno escluse dall'analisi. Si assume che i valori inclusi nell'intervallo siano interi (i non interi verranno troncati).

Risposte multiple, tabelle di contingenza: Opzioni

Percentuali delle celle. I conteggi delle celle vengono visualizzi sempre. È possibile impostare la visualizzazione di percentuali di riga, percentuali di colonna e percentuali per tabelle a due vie (totale).

Percentuali basate su. È possibile basare le percentuali delle celle sui casi (o rispondenti). Questa opzione non è disponibile se si seleziona la corrispondenza delle variabili tra insiemi a categorie multiple. È inoltre possibile basare le percentuali delle celle sulle risposte. Per gli insiemi a dicotomie multiple, il numero di risposte equivale al numero di valori conteggiati nei diversi casi. Per gli insiemi a categorie multiple, il numero di risposte equivale al numero di valori dell'intervallo definito.

Valori mancanti. È possibile scegliere una delle seguenti opzioni o entrambe:

- **Escludi casi listwise all'interno delle dicotomie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabulazione dell'insieme a dicotomie multiple. Questa opzione può essere applicata solo agli insiemi a risposta multipla definiti come insiemi dicotomici. Per impostazione predefinita, un caso viene considerato mancante per un insieme a dicotomie multiple se nessuna delle variabili che lo compongono contiene il valore conteggiato. I casi con valori mancanti solo per alcune variabili verranno inclusi nelle tabulazioni del gruppo se almeno una variabile contiene il valore conteggiato.
- **Escludi casi listwise all'interno delle categorie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabulazione dell'insieme a categorie multiple. Questa opzione viene applicata solo a insiemi a risposta multipla definiti come insiemi di categorie. Per impostazione predefinita, un caso viene considerato mancante per un insieme a categorie multiple solo se nessuno dei componenti contiene valori validi all'interno dell'intervallo definito.

Per impostazione predefinita, quando si incrociano due insiemi a categorie multiple, ciascuna variabile del primo gruppo verrà ordinata in tabelle con ciascuna variabile del secondo gruppo e quindi verranno sommati i conteggi relativi a ciascuna cella. Alcune risposte, pertanto, potranno comparire più volte nella stessa tabella. È possibile scegliere la seguente opzione:

Metti in corrispondenza le variabili tra gli insiemi di risposte. Consente di associare la prima variabile del primo gruppo con la prima variabile del secondo gruppo e così via. Se viene selezionata questa opzione, le percentuali delle celle saranno basate sulle risposte e non sui rispondenti. Questa opzione non è disponibile per gli insiemi a dicotomie multiple né per le variabili elementari.

Funzioni aggiuntive del comando MULT RESPONSE

Il linguaggio della sintassi dei comandi consente inoltre di:

- Ottenere tabelle delle tavole di contingenza con un massimo di cinque dimensioni (con il sottocomando BY).
- Modificare le opzioni di formattazione dell'output, inclusa l'eliminazione delle etichette valore (con il sottocomando FORMAT).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Risultati di report

Risultati di report

Gli elenchi dei casi e le statistiche descrittive sono strumenti fondamentali per lo studio e la presentazione dei dati. Per creare elenchi dei casi è possibile utilizzare l'Editor dei dati o la procedura Riassumi, per produrre conteggi della frequenza e statistiche descrittive è possibile utilizzare la procedura Frequenze, mentre per creare statistiche per la sottopopolazione è possibile utilizzare la procedura Medie. Queste procedure utilizzano un formato progettato per rendere chiare le informazioni. Per visualizzare le informazioni in un formato diverso, è possibile impostare la presentazione dei dati mediante le opzioni per i report Riepiloghi per righe e Riepiloghi per colonne.

Report : Riepiloghi per righe

La procedura Report: Riepiloghi per righe consente di creare report in cui statistiche di riepilogo diverse sono disposte in righe distinte. Sono inoltre disponibili gli elenchi dei casi, che possono includere o meno le statistiche di riepilogo.

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

Esempio. Una ditta proprietaria di una catena di negozi al dettaglio registra le informazioni sui dipendenti, che includono informazioni sugli stipendi e le mansioni nonché sul negozio e il reparto in cui lavora ogni dipendente. È quindi possibile creare un report che includa le informazioni relative a ogni impiegato (elenco) suddivise per negozio e per reparto (variabili di interruzione) e che includa statistiche di riepilogo (ad esempio, lo stipendio medio) per ciascun negozio e reparto nonché per ciascun reparto di ogni negozio.

Colonne dati. Elenca le variabili da rappresentare nel report, per le quali si desidera creare elenchi dei casi o statistiche di riepilogo e controlla il formato per la visualizzazione delle colonne di dati.

Colonne di interruzione. Elenca le variabili di interruzione facoltative che suddividono il report in più gruppi e consente di gestire le statistiche di riepilogo e i formati per la visualizzazione delle colonne di interruzione. Se sono presenti più variabili di interruzione, per ogni categoria di ciascuna variabile di interruzione verrà creato un gruppo distinto all'interno delle categorie della variabile di interruzione precedente nell'elenco. Le variabili di interruzione devono essere variabili categoriali discrete che suddividono i casi in un numero limitato di categorie significative. I singoli valori di ciascuna variabile di interruzione vengono visualizzati, ordinati, in una colonna distinta a sinistra di tutte le colonne dati.

Report. Consente di controllare le caratteristiche generali del report, inclusi i titoli, le statistiche di riepilogo globali, la visualizzazione dei valori mancanti e la numerazione delle pagine.

Visualizza casi. Consente di visualizzare i valori effettivi (o etichette valore) delle variabili delle colonne di dati per ciascun caso. In tal modo viene creato un elenco che potrebbe risultare molto più lungo di un report di riepilogo.

Anteprima. Consente di visualizzare solo la prima pagina del report. Questa opzione è utile per visualizzare in anteprima il formato del report prima di generarlo.

I dati sono già ordinati. Se il report include variabili di interruzione, prima di generarlo è necessario ordinare il file di dati in base ai valori delle variabili di interruzione. Se il file di dati è già ordinato in base ai valori delle variabili di interruzione, è possibile ridurre il tempo di elaborazione selezionando questa opzione. Questa opzione risulta particolarmente utile dopo aver visualizzato un'anteprima del report.

Ottenere un report di riepilogo: riepiloghi per righe

1. Dai menu, scegliere:

   Analizza > **Report** > **Report : Riepiloghi per righe...**
2. Selezionare una o più variabili per Colonne dati. Per ogni variabile selezionata verrà generata una colonna nel report.
3. Per i report ordinati e visualizzati in base ai sottogruppi, selezionare una o più variabili per Colonne di interruzione.
4. Per i report con statistiche di riepilogo per i sottogruppi definiti in base alle variabili di interruzione, selezionare la variabile di interruzione nell'elenco Variabili di colonna di interruzione e fare clic su **Riepilogo** nel gruppo Colonne di interruzione per specificare le misure di riepilogo.
5. Per i report con statistiche di riepilogo globali, fare clic su **Riepilogo** per specificare le misure di riepilogo.

Formato delle colonne e di interruzione del report

Nelle finestre di dialogo relative al formato è possibile impostare i titoli e la larghezza delle colonne, l'allineamento del testo e la visualizzazione dei valori dei dati o delle etichette valore. L'opzione Formato colonna dati consente di impostare il formato delle colonne dati nella parte destra della pagina del report. L'opzione Formato di interruzione consente di impostare il formato delle colonne di interruzione nella parte sinistra.

Titolo della colonna. Consente di impostare il titolo della colonna per la variabile selezionata. I titoli lunghi vanno a capo automaticamente all'interno della colonna. Per inserire manualmente le interruzioni di riga nella posizione in cui si desidera che i titoli vadano a capo, è possibile utilizzare il tasto Invio.

Posizione valore nella colonna. Per la variabile selezionata, consente di impostare l'allineamento dei valori o delle etichette valore all'interno della colonna. L'allineamento dei valori o delle etichette non modifica l'allineamento delle intestazioni di colonna. È possibile rientrare il contenuto delle colonne di un numero specifico di caratteri oppure centrarlo.

Contenuto della colonna. Per la variabile selezionata, consente di impostare la visualizzazione dei valori dei dati o delle etichette valore definite. I valori dei dati per i quali non è stata definita alcuna etichetta vengono sempre visualizzati. Non è disponibile per le colonne dati nei report di riepilogo per colonne.

Report: Righe di riepilogo per/Righe di riepilogo finali

Le due finestre di dialogo Report: Righe di riepilogo consentono di impostare la visualizzazione della statistica di riepilogo per i gruppi di interruzione e per l'intero report. Righe di riepilogo consente di impostare la statistica del sottogruppo per ciascuna categoria definita tramite le variabili di interruzione. Righe di riepilogo finali consente di impostare le statistiche globali visualizzate nella parte finale del report.

Le statistiche di riepilogo disponibili sono: somma, media, minimo, massimo, numero di casi, percentuale di casi al di sopra o al di sotto di un valore specifico, percentuale di casi entro un intervallo specifico di valori, deviazione standard, curtosi, varianza e asimmetria.

Report: Opzioni di interruzione

La funzione Opzioni di interruzione consente di impostare la spaziatura e l'impaginazione delle informazioni sui gruppi.

Controllo pagina. Consente di impostare la spaziatura e l'impaginazione per le categorie relative alla variabile di interruzione selezionata. È possibile impostare il numero desiderato di righe vuote tra i gruppi o fare in modo che ciascun gruppo inizi in una nuova pagina.

Righe vuote prima dei riepiloghi. Consente di impostare il numero delle linee vuote tra le etichette dell'asse della categoria o i dati e le statistiche di riepilogo. Questa funzionalità risulta particolarmente utile per i report combinati che includono sia l'elenco dei singoli casi che le statistiche di riepilogo per i

gruppi. In questo tipo di report è possibile inserire una spaziatura tra l'elenco dei casi e le statistiche di riepilogo.

Report: Opzioni

La funzione Report: Opzioni consente di impostare la modalità di elaborazione e la visualizzazione dei valori mancanti e la numerazione delle pagine del report.

Escludi casi con valori mancanti listwise . Consente di eliminare dal report i casi con valori mancanti per le variabili del report.

Valori mancanti visualizzati come . Consente di specificare il simbolo che rappresenta i valori mancanti nel file di dati. Il simbolo deve essere costituito da un solo carattere e viene usato per rappresentare sia i *valori mancanti di sistema* che i *valori mancanti definiti dall'utente*.

Numero di pagine da. Consente di specificare un numero di pagina con cui contrassegnare la prima pagina del report.

Report: Layout

L'opzione Report: Layout consente di impostare la larghezza e la lunghezza di ogni pagina del report, la posizione del report nella pagina e l'inserimento di linee vuote ed etichette.

Layout di pagina. Consente di impostare i margini della pagina espressi in linee (superiori e inferiori) e caratteri (a destra e a sinistra) nonché l'allineamento dei report all'interno dei margini.

Titoli e piè di pagina. Consente di impostare il numero delle linee che separano i piè di pagina e i titoli della pagina dal corpo del report.

Colonne di interruzione. Consente di impostare la visualizzazione delle colonne di interruzione. Se vengono specificate più variabili di interruzione, queste possono trovarsi in colonne diverse oppure nella prima colonna. Se tutte le variabili di interruzione vengono inserite nella prima colonna, verrà creato un report di larghezza minore.

Titoli di colonna. Consente di impostare la visualizzazione dei titoli di colonna, inclusi lo spazio tra i titoli e il corpo del report, la sottolineatura del titolo e l'allineamento verticale dei titoli di colonna.

Righe colonna di dati ed etichette di interruzione. Consente di posizionare le informazioni relative alle colonne dati (valori dei dati e/o statistiche di riepilogo) in relazione alle etichette di interruzione all'inizio di ogni gruppo. La prima riga delle informazioni sulla colonna dati può iniziare sulla stessa riga in cui si trova l'etichetta di gruppo o dopo un numero specifico di righe rispetto alla posizione dell'etichetta di gruppo. Non è disponibile per i report di riepilogo per colonne.

Report: Titoli

L'opzione Report: Titoli consente di impostare il contenuto e la posizione dei titoli e dei piè di pagina dei report. È possibile specificare fino a dieci righe per i titoli di pagina e per i piè di pagina, con componenti centrati oppure allineati a destra o a sinistra in ciascuna riga.

Se si inseriscono variabili in titoli e in piè di pagina, l'etichetta valore corrente o il valore della variabile verrà visualizzato nel titolo o nel piè di pagina. Nei titoli viene visualizzata l'etichetta valore corrispondente al valore della variabile all'inizio della pagina. Nei piè di pagina viene visualizzata l'etichetta valore corrispondente al valore della variabile al termine della pagina. Se non sono presenti etichette valore, viene visualizzato il valore effettivo.

Variabili speciali. Le variabili speciali *DATE* e *PAGE* consentono di inserire la data corrente o il numero di pagina in una delle righe dell'intestazione o del piè di pagina del report. Se il file di dati utilizzato contiene le variabili denominate *DATE* o *PAGE*, non sarà possibile utilizzare tali variabili nei titoli o nei piè di pagina.

Report: Riepiloghi per colonne

L'opzione Report: Riepiloghi per colonne consente di creare report di riepilogo in cui le statistiche di riepilogo vengono visualizzate in colonne distinte.

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

Esempio. Un'azienda proprietaria di una catena di negozi al dettaglio conserva le registrazioni delle informazioni sui dipendenti, tra cui gli stipendi, le mansioni e il reparto in cui lavora ogni dipendente. È quindi possibile creare un report che includa le statistiche di riepilogo sugli stipendi (ad esempio media, minimo e massimo) per ogni reparto.

Colonne dati. Consente di visualizzare un elenco delle variabili da rappresentare nel report e per le quali si desidera produrre statistiche di riepilogo e di impostare il formato di visualizzazione e le statistiche di riepilogo visualizzate per ogni variabile.

Colonne di interruzione. Consente di visualizzare l'elenco delle variabili di interruzione facoltative che suddividono il report in più gruppi e di impostare i formati di visualizzazione delle colonne di interruzione. Se sono presenti più variabili di interruzione, per ogni categoria di ciascuna variabile di interruzione verrà creato un gruppo distinto all'interno delle categorie della variabile di interruzione precedente nell'elenco. Le variabili di interruzione devono essere variabili categoriali discrete che suddividono i casi in un numero limitato di categorie significative.

Report. Consente di impostare le caratteristiche globali del report, inclusi i titoli, la visualizzazione dei valori mancanti e la numerazione delle pagine.

Anteprima. Consente di visualizzare solo la prima pagina del report. Questa opzione è utile per visualizzare in anteprima il formato del report prima di generarlo.

I dati sono già ordinati. Se il report include variabili di interruzione, prima di generarlo è necessario ordinare il file di dati in base ai valori delle variabili di interruzione. Se il file di dati è già ordinato in base ai valori delle variabili di interruzione, è possibile ridurre il tempo di elaborazione selezionando questa opzione. Questa opzione risulta particolarmente utile dopo aver visualizzato un’anteprima del report.

Ottenere un report di riepilogo: riepiloghi per colonne

1. Dai menu, scegliere:

   Analizza > **Report** > **Report: Riepiloghi per colonne...**
2. Selezionare una o più variabili per Colonne dati. Per ogni variabile selezionata verrà generata una colonna nel report.
3. Per modificare la misura di riepilogo relativa a una variabile, selezionare la variabile nell'elenco Variabili di colonna di dati e fare clic su **Riepilogo**.
4. Per ottenere più di una misura di riepilogo per una variabile, selezionare la variabile nell'elenco origine e spostarla nell'elenco Variabili di colonna di dati più volte, una per ogni misura di riepilogo desiderata.
5. Per visualizzare una colonna contenente la somma, la media, il rapporto o altre funzioni per le colonne esistenti, fare clic su **Inserisci totale**. In tal modo verrà inserita una variabile denominata *totale* nell'elenco Colonne dati.
6. Per i report ordinati e visualizzati in base ai sottogruppi, selezionare una o più variabili per Colonne di interruzione.

Funzione di riepilogo delle colonne di dati

L'opzione Righe di riepilogo consente di controllare la statistica di riepilogo visualizzata per la variabile della colonna dati selezionata.

Le statistiche di riepilogo disponibili sono: somma, media, minimo, massimo, numero di casi, percentuale di casi al di sopra o al di sotto di un valore specifico, percentuale di casi entro un intervallo specifico di valori, deviazione standard, varianza, curtosi e asimmetria.

Colonna di riepilogo del totale generale

L'opzione Colonna di riepilogo consente di gestire le statistiche di riepilogo generali che riassumono due o più colonne dati.

Le statistiche di riepilogo generali disponibili sono somma di colonne, media di colonne, minimo, massimo, differenza tra valori in due colonne, quoziente dei valori in una colonna divisi per i valori in un'altra colonna e prodotto di valori di colonne moltiplicati insieme.

Somma di colonne . La colonna *totale* rappresenta la somma delle colonne nell'elenco Colonna di riepilogo.

Media di colonne. La colonna *totale* rappresenta la media delle colonne nell'elenco Colonna di riepilogo.

Minimo di colonne. La colonna *totale* rappresenta il minimo delle colonne nell'elenco Colonna di riepilogo.

Massimo di colonne. La colonna *totale* rappresenta il massimo delle colonne nell'elenco Colonna di riepilogo.

1° colonna - 2° colonna. La colonna *totale* rappresenta la differenza delle colonne nell'elenco Colonna di riepilogo. L'elenco Colonna di riepilogo deve contenere esattamente due colonne.

1° colonna / 2° colonna. La colonna *totale* rappresenta il quoziente delle colonne nell'elenco Colonna di riepilogo. L'elenco Colonna di riepilogo deve contenere esattamente due colonne.

% 1° colonna / 2° colonna. La colonna *totale* rappresenta la percentuale della prima colonna rispetto alla seconda colonna nell'elenco Colonna di riepilogo. L'elenco Colonna di riepilogo deve contenere esattamente due colonne.

Prodotto di colonne. La colonna *totale* rappresenta il prodotto delle colonne nell'elenco Colonna di riepilogo.

Formato delle colonne del report

Le opzioni di formattazione delle colonne dati e di interruzione per i report di riepilogo per colonne sono le stesse descritte per i report di riepilogo per righe.

Report: Opzioni di interruzione (Riepiloghi per colonne)

La funzione Opzioni di interruzione consente di impostare la visualizzazione, la spaziatura e l'impaginazione per i gruppi.

Totale parziale. Consente di impostare la visualizzazione di totali parziali per i gruppi.

Controllo pagina. Consente di impostare la spaziatura e l'impaginazione per le categorie relative alla variabile di interruzione selezionata. È possibile impostare il numero desiderato di righe vuote tra i gruppi o fare in modo che ciascun gruppo inizi in una nuova pagina.

Righe vuote prima del totale parziale. Consente di impostare il numero di righe vuote tra i dati dei gruppi e i totali parziali.

Report: Opzioni (Riepiloghi per colonne)

Le opzioni consentono di impostare la visualizzazione dei totali generali e dei valori mancanti e l'impaginazione dei report di riepilogo per colonne.

Totale finale. Consente di visualizzare ed etichettare un totale finale per ogni colonna, visualizzato alla fine della colonna.

Valori mancanti. È possibile escludere i valori mancanti dal report o selezionare un solo carattere per indicare i valori mancanti nel report.

Report: Layout per riepiloghi per colonne

Le opzioni di layout per i report di riepilogo per colonne sono le stesse descritte per i report di riepilogo per righe.

Funzioni aggiuntive del comando REPORT

Il linguaggio della sintassi dei comandi consente inoltre di:

- Visualizzare diverse funzioni di riepilogo nelle colonne di una riga di riepilogo.
- Inserire righe di riepilogo nelle colonne dati per le variabili diverse dalle variabili della colonna, o per le varie combinazioni (funzioni composte) di funzioni di riepilogo.
- Utilizzare Mediana, Moda, Frequenza e Percentuale come funzioni di riepilogo.
- Controllare in modo più preciso il formato di visualizzazione delle statistiche di riepilogo.
- Inserire linee vuote in corrispondenza di vari punti nei report.
- Inserire linee vuote dopo ogni *n*° caso nei report elencati.

A causa della complessità della sintassi REPORT, può risultare utile, durante la costruzione di un nuovo report con la sintassi, approssimare il report generato dalle finestre di dialogo, copiare e incollare la sintassi corrispondente e ridefinire tale sintassi in modo da farla corrispondere al report specifico.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Analisi di affidabilità

L'analisi di affidabilità consente di studiare le proprietà delle scale di misurazione e degli elementi che le compongono. La procedura Analisi di affidabilità calcola una serie di misure comunemente utilizzate in relazione all'affidabilità della scala e fornisce inoltre informazioni relative alle relazioni tra singoli elementi della scala. I coefficienti di correlazione intra - class possono essere utilizzati per calcolare le stime di affidabilità inter - rater.

L'analisi di affidabilità fornisce anche le statistiche Kappa multirater di Fleiss che valutano l'accordo interrater per determinare l'affidabilità tra i vari stimatori. Un accordo più alto garantisce maggiore fiducia negli ascolti che riflettono la vera circostanza. Le opzioni Fleiss ' Multiple Rater Kappa sono disponibili nella finestra di dialogo "Analisi di affidabilità: Statistiche" a pagina 264 .

Esempio
Il questionario misura la soddisfazione del cliente in un modo utile? Utilizzando l'analisi di affidabilità, è possibile determinare il grado di correlazione tra gli elementi del questionario, ottenere un indice globale della ripetibilità oppure la concordanza interna della scala in modo globale. È quindi possibile identificare gli elementi del problema che devono essere esclusi dalla scala.

Statistiche
Descrittivi per ogni variabile e per la scala, statistiche di riepilogo su voci, correlazioni interitem e covarianze, stime di affidabilità, tabella ANOVA, coefficienti di correlazione intraclasse, T^2, Test di additività di Tukey, e Fleiss ' Multiple Rater Kappa.

Modelli
Sono disponibili i seguenti modelli di affidabilità:

Alfa (Cronbach)
Questo modello è una misura della coerenza interna in base alla correlazione interelemento media.

Omega (McDonald's)

Questo modello presuppone che il modello sia unidimensionale incluso un singolo fattore senza alcuna dipendenza da voce locale sotto forma di covarianze di errore. Il modello implica che la covarianza delle due diverse voci sia il prodotto dei loro saccheggi.

Suddivisione a metà
Questo modello suddivide la scala in due parti ed esamina la correlazione tra le parti.

Guttman
Questo modello calcola i limiti inferiori di Guttman per una reale affidabilità.

Parallelo
Questo modello presume che tutti gli elementi abbiano varianze e varianze di errore uguali tra le repliche.

Parallelo esatto
Questo modello afferma le ipotesi del modello parallelo e assume inoltre medie uguali degli elementi.

Considerazioni sui dati dell'analisi di affidabilità

Dati
I dati possono essere dicotomici, ordinali oppure intervalli ma devono essere codificati numericamente.

Ipotesi
Le osservazioni devono essere indipendenti e gli errori non devono essere correlati agli elementi. Ogni coppia di elementi deve avere una distribuzione normale bivariata. Le scale devono essere additive, in modo che ogni elemento sia correlato in modo lineare al punteggio totale. Le seguenti ipotesi si applicano per le statistiche Fleiss ' Multiple Rater Kappa:

- Almeno due variabili di item devono essere selezionate per eseguire qualsiasi statistica di affidabilità.
- Quando vengono selezionate almeno due variabili di rating, viene incollata la sintassi Fleiss ' Multiple Rater Kappa.
- Non ci sono connessioni tra gli stimatori.
- Il numero di stimatori è una costante.
- Ogni soggetto è valutato dallo stesso gruppo contenente solo un solo rater.
- Nessun peso può essere assegnato ai vari disaccordi.

Procedure correlate
Se si desidera esplorare la dimensionalità degli elementi della scala (per verificare se è necessaria più di una costruzione per tenere conto del modello dei punteggi degli elementi), utilizzare la procedura Analisi fattoriale o Scaling multidimensionale. Per identificare i gruppi omogenei di variabili, usare l'analisi cluster gerarchica per raggruppare le variabili.

Per ottenere l'analisi di affidabilità

1. Dai menu, scegliere:

   Analizza > **Scala** > **Analisi di affidabilità...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare due o più variabili come potenziali componenti di una scala additiva.
3. Scegli un modello dall'elenco a discesa **Modello** .
4. Opzionalmente, clicca su **Statistiche** per selezionare varie statistiche che descrivono gli articoli di scala o l'accordo di interruttore.

Analisi di affidabilità: Statistiche

È possibile selezionare varie statistiche che descrivono la tua scala, gli articoli e l'accordo di interruttore per determinare l'affidabilità tra i vari rateri. Le statistiche riportate per impostazione predefinita comprendono il numero di casi, il numero di elementi e le stime di affidabilità riportati di seguito:

Modelli alfa
Per i dati dicotomici, è equivalente al coefficiente Kuder-Richardson 20 (KR20).

Modelli omega
Stima di McDonald’s omega per valutare l'affidabilità.

Modelli suddivisione a metà
Correlazione tra stime dei parametri, affidabilità di Suddivisione a metà di Guttman, affidabilità di Spearman-Brown (lunghezza uguale e diversa) e coefficiente alfa per ogni metà.

Modelli di Guttman
Coefficienti di affidabilità da lambda 1 a lambda 6.

Modelli paralleli e Strict paralleli
Test sulla bontà dell'adattamento del modello, stime della varianza di errore, varianza comune e true, correlazione interelemento comune stimata, affidabilità stimata e stima di affidabilità non distorta.

Descrittive per
Fornisce statistiche descrittive per le scale o per gli elementi tra i casi.

Elemento
Fornisce statistiche descrittive per gli elementi tra i casi.

Scala
Fornisce statistiche descrittive per le scale.

Scala se l'elemento è eliminato
Consente di visualizzare statistiche di riepilogo per il confronto di ogni elemento con la scala composta dagli altri elementi. Le statistiche includono la media della scala e la varianza risultante se l'elemento venisse eliminato dalla scala, la correlazione tra l'elemento e la scala composta dagli altri elementi e l'Alfa di Cronbach risultante se l'elemento venisse eliminato dalla scala.

Riepiloghi
Fornisce statistiche descrittive della distribuzione di elementi tra tutti gli elementi nella scala.

Medie
Statistiche di riepilogo per le medie degli elementi. Vengono visualizzate la media degli elementi più piccola, la più grande e la media, nonché l'intervallo e la varianza standard delle medie degli elementi e il rapporto fra la media degli elementi più grande e quella più piccola.

Varianze
Statistiche di riepilogo per le varianze degli elementi. Vengono riprodotte la varianza degli elementi minima, massima e media, l'intervallo e la varianza delle varianze degli elementi e il rapporto fra la varianza degli elementi più grande e quella più piccola.

Correlazioni
Statistiche di riepilogo per le correlazioni tra elementi. Vengono visualizzate le correlazioni interelemento più piccola, più grande e media, l'intervallo e la varianza delle correlazioni interelemento e il rapporto tra le correlazioni interelemento più grandi e più piccole.

Covarianze
Statistiche di riepilogo per le covarianze tra elementi. Vengono visualizzate le covarianze interelemento più piccola, più grande e media, l'intervallo e la varianza delle covarianze interelemento e il rapporto tra le covarianze interelemento più grandi e più piccole.

Interelemento
Fornisce matrici di correlazioni o covarianze tra elementi.

Tabella ANOVA
Fornisce test di medie uguali.

Test F
Visualizza una tabella di analisi della varianza a misure ripetute.

Chi-quadrato di Friedman
Visualizza il chi - quadrato di Friedman e il coefficiente di concordanza di Kendall. Questa opzione è appropriata per dati che rappresentano classifiche (ranghi). Il test del chi-quadrato sostituisce il test F solitamente usato nella tabella ANOVA.

Chi-quadrato di Cochran
Visualizza la Q di Cochran. Questa opzione è appropriata per i dati che sono dicotomici. La statistica Q sostituisce la statistica F solitamente usata nella tabella ANOVA.

Accordo interrater: Kappa di Fleiss

Valuta l'accordo di interruttore per determinare l'affidabilità tra i vari rateri. Un accordo più alto garantisce maggiore fiducia negli ascolti che riflettono la vera circostanza. La statistica kappa non ponderata generalizzata misura l'accordo tra qualsiasi numero costante di rateri mentre assume:

- Devono essere specificate almeno due variabili di item per eseguire qualsiasi statistica di affidabilità.
- Devono essere specificate almeno due variabili di punteggio.
- Le variabili selezionate come voci possono essere selezionate anche come ascolti.
- Non ci sono connessioni tra gli stimatori.
- Il numero di stimatori è una costante.
- Ogni soggetto è valutato dallo stesso gruppo contenente solo un solo rater.
- Nessun peso può essere assegnato ai vari disaccordi.

Visualizza accordo su singole categorie

Specifica se emettere o meno l'accordo sulle singole categorie. Per impostazione predefinita, l'output elimina la stima su qualsiasi singola categoria. Quando è abilitata, più tabelle vengono visualizzate nell'output.

Ignora casi stringa

Controlla se le variabili di stringa sono sensibili al caso. Per impostazione predefinita, i valori di abilitazione stringa sono sensibili al caso.

Le etichette di categoria della stringa vengono visualizzate in caratteri maiuscoli

Controlla se le etichette di categoria nelle tabelle di output vengono visualizzate in maiuscolo o in minuscolo. L'impostazione è abilitata per impostazione predefinita, che visualizza in maiuscolo le etichette di categoria stringa.

Livello di significatività asintotica (%)

Specifica il livello di significatività per gli intervalli di confidenza asintotici. 95 è l'impostazione predefinita.

Mancante/i

Escludi valori mancanti di sistema e definiti dall'utente

Controlla l'esclusione dei valori mancanti e mancanti di sistema. Per impostazione predefinita, sono esclusi i valori mancanti dell'utente e del sistema.

I valori mancanti definiti dall'utente vengono considerati come validi

Quando abilitato, tratta gli utenti - mancanti e i valori mancanti di sistema come dati validi. L'impostazione è disabilitata per impostazione predefinita.

T-quadrato di Hotelling

Crea un test multivariato dell'ipotesi nulla in base alla quale tutti gli elementi sulla scala hanno la stessa media.

Test di additività di Tukey

Produce un test del presupposto che non esista un'interazione moltiplicativa tra gli elementi.

Coefficiente di correlazione intraclasse

Crea misurazioni della consistenza o dell'accordo dei valori all'interno dei casi.

Modello

Consente di selezionare il modello per calcolare i coefficienti di correlazione intraclasse. I modelli disponibili sono A due vie misto, A due vie casuale e A una via casuale. Selezionare **A due vie misto** quando gli effetti relativi alle persone sono casuali e quelli relativi all'elemento sono fissi, **A due vie casuale** quando sia gli effetti relativi alle persone che quelli relativi all'elemento sono casuali oppure **A una via casuale** quando gli effetti relativi alle persone sono casuali.

Tipo

Consente di selezionare il tipo di indice. I tipi disponibili sono Uniformità e Accordo assoluto.

Intervallo di confidenza (%)

Consente di specificare il livello relativo all'intervallo di confidenza. Il valore predefinito è il 95%.

Valore test
Consente di specificare il valore ipotizzato del coefficiente relativo al test sull'ipotesi. Si tratta del valore rispetto al quale viene confrontato il valore osservato. Il valore predefinito è 0.

Funzioni aggiuntive del comando RELIABILITY

Il linguaggio della sintassi dei comandi consente inoltre di:

- Leggere ed analizzare una matrice di correlazione.
- Scrivere una matrice di correlazione per analizzarla in seguito.
- Specificare le suddivisioni diverse dalle metà uguali per il metodo di suddivisione a metà.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Kappa pesata

La kappa ponderata di Cohen è ampiamente utilizzata nella classificazione incrociata come misura d'intesa tra i rateri osservati. Si tratta di un indice di accordo adeguato quando gli ascolti sono scale nominali senza struttura d'ordine. Lo sviluppo della kappa ponderata di Cohen era motivato dal fattore che alcune assegnazioni in un tavolo di contingenza potrebbero essere di maggiore gravità rispetto agli altri. La statistica si affida ai pesi cellulari predefiniti che riflettono accordo o disaccordo.

La Procedura Ponderata Kappa fornisce opzioni per stimare la kappa ponderata di Cohen, un'importante generalizzazione della statistica kappa che misura l'accordo di due soggetti ordinali con categorie identiche.

Nota: La Procedura di Kappa ponderata sostituisce le funzionalità precedentemente fornite dall'estensione `STATS WEIGHTED KAPPA.spe`.

Esempio
Ci sono situazioni in cui le differenze tra i rateri non vanno trattate altrettanto importanti. Un esempio sarebbe nell'industria sanitaria dove più persone raccolgono ricerche o dati clinici. In questi casi l'attendibilità dei dati può entrare in discussione data la variabilità tra quelli che raccolgono i dati.

Statistiche
La kappa ponderata di Cohen, scala lineare, scala quadratica, intervallo di confidenza asintotico.

Considerazioni sui dati Kappa ponderati

Dati
È necessaria una tabella a due vie che si basa su un dataset attivo al fine di stimare la statistica kappa ponderata di Cohen.

Le variabili di abilitazione devono essere dello stesso tipo (tutte le stringhe o tutte numeriche).

La stima della kappa ponderata di Cohen ha senso solo quando le categorie delle due variabili di rating, rappresentate dalla riga e dalla colonna nella tabella, sono opportunamente ordinate (per una coppia di variabili numeriche viene applicata l'ordine numerico; per una coppia di variabili stringa viene applicato l'ordine alfabetico).

Ipotesi
Quando vengono selezionate coppie di variabili miste, la kappa ponderata di Cohen non è stimata.

Le variabili di abilitazione sono assunte per condividere lo stesso insieme di categorie.

Per ottenere un'analisi Kappa ponderata

1. Dai menu, scegliere:

   Analizzare > **Scala** > **Kappa Pesata ...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare due o più variabili o variabili numeriche da specificare come **rateri Pairwise**.

   Nota: È necessario selezionare tutte le variabili stringa o tutte le variabili numeriche.

3. Opzionalmente, abilitare la impostazione **Specificare i livelli per le righe e le colonne** per controllare la visualizzazione di rateri pairwise o di righe / colonne.

   - Quando è abilitata, i rateri pairwise vengono soppressi e i livelli di riga / colonna sono visualizzati. L'interfaccia utente si aggiorna per fornire i campi **Rater (s)** e **Rater di Colonna** (sostituendo in modo efficace il campo **Pairwise raters** .
   - Quando i disabili, i rater row/column sono soppressi e i livelli di rater pairwise (impostazione predefinita)

   Quando **Specificare i livelli per righe e colonne** è abilitato, specificare almeno una variabile per entrambi **Rater di riga** e **Rater di Colonna**.

   Nota: Se entrambi **Rater di riga** e **Rater di Colonna** contengono solo una variabile, le variabili selezionate non possono essere uguali per entrambi.

4. Opzionalmente, clicca su **Criteri** per specificare la scala di ponderazione e le impostazioni dei valori mancanti o **Stampa** per specificare il formato di visualizzazione e le impostazioni di contingenza.

Kappa pesata: criteri

Dialogo Criteri fornisce opzioni per specificare la stima delle statistiche kappa ponderate di Cohen.

Scala di peso
Fornisce opzioni per specificare una ponderazione lineare o quadratica per l'accordo. L'uso del peso lineare (noto anche come Cicchetti-Allison) è l'impostazione predefinita.

Valori mancanti
Fornisce opzioni per la rimozione dei casi con valori mancanti su base pairwise e per il trattamento dei valori mancanti degli utenti come valido.

Le variabili di classificazione stringa sono sensibili al maiuscolo/minuscolo
Quando selezionate, le variabili di stringa vengono trattate come sensibili al caso.

Intervallo di confidenza asintotico (%)
Questa impostazione opzionale specifica il livello di confidenza per la stima degli intervalli di confidenza asintotici. Deve essere un singolo doppio valore compreso tra 0 e 100 (95 è l'impostazione predefinita).

Kappa pesata: stampa

La finestra di dialogo Stampa fornisce le opzioni per il controllo delle tabelle di contingenza.

Visualizzazione e formato
Fornisce le opzioni per il controllo del display e del formato della tabella di stabulazione.

Le categorie di classificazione sono visualizzate in ordine crescente
Quando selezionate, le categorie di abilitazione nelle tabelle crosstabulari visualizzano in ordine crescente. Quando l'impostazione non viene selezionata, le categorie di abilitazione nelle tabelle crosiche visualizzano in ordine decrescente. L'impostazione è abilitata per impostazione predefinita.

Le etichette di categoria della stringa vengono visualizzate in caratteri maiuscoli
Quando selezionato, i tabelle di contingenza vengono visualizzati come maiuscole. Quando l'impostazione non è selezionata, i tavoli di contingenza vengono visualizzati come minuscoli. L'impostazione è abilitata per impostazione predefinita.

Tavola di contingenza
Fornisce le opzioni per specificare le variabili di abilitazione utilizzate nella crosstabulazione. Per impostazione predefinita, le impostazioni di contingenza non sono abilitate, che sopprime la crosstabulazione di eventuali variabili di rating.

Visualizzare la tavola di contingenza delle variabili di classificazione
Quando selezionata, questa impostazione consente la crosstabulazione di tutte o le variabili di abilitazione specificate dall'utente.

Includere tutte le variabili di classificazione
Quando selezionate, le tabelle di contingenza vengono stampate per tutte le coppie di variabili di rating definite.

Includere variabili di classificazione specificate dall'utente
Quando selezionate, utilizzare le **Variabili disponibili**, **Rater di riga (s)**e i campi **Rater di Colonna** per selezionare quali variabili di abilitazione sono incluse nelle tabelle di contingenza.

Scaling multidimensionale

La procedura Scaling multidimensionale consente di effettuare un tentativo per trovare la struttura in un insieme di misure della distanza tra oggetti o casi. Questa attività viene compiuta assegnando le osservazioni a ubicazioni specifiche in uno spazio concettuale (in genere bi o tridimensionale) in modo che le distanze tra i punti nello spazio corrispondano il più possibile alle dissimilarità specificate. In molti casi, le dimensioni di questo spazio concettuale possono essere interpretate ed utilizzate allo scopo di comprendere meglio i dati.

Se si dispone di variabili misurate oggettivamente, è possibile utilizzare lo scaling multidimensionale come una tecnica di riduzione dei dati (la procedura calcolerà le distanze dai dati multivariati per l'utente, se necessario). È inoltre possibile applicare lo scaling multidimensionale ai punteggi soggettivi di dissimilarità tra oggetti o concetti. In aggiunta, Scaling multidimensionale può gestire i dati di dissimilarità da più origini, come nel caso di più stimatori o di rispondenti ai questionari.

Esempio. In che modo i consumatori percepiscono le similitudini tra automobili diverse? Se si dispone di dati che rilevano punteggi di similarità tra diverse forme e modelli di automobili, lo scaling multidimensionale consentirà di identificare le dimensioni in grado di descrivere le percezioni dei consumatori. È possibile verificare, ad esempio, che il prezzo e le dimensioni di un veicolo definiscono uno spazio bidimensionale, secondo le spiegazioni fornite dai rispondenti.

Statistiche. Per ogni modello: matrice dei dati, matrice dei dati scalati in modo ottimale, S-stress (di Young), stress (di Kruskal), RSQ, coordinate degli stimoli, stress medio e RSQ per ogni stimolo (modelli RMDS). Per i modelli delle differenze singole (INDSCAL): pesi del soggetto e indice di stranezza per ogni soggetto. Per ogni matrice nei modelli di scaling multidimensionale replicati: stress e RSQ per ogni stimolo. Grafici: coordinate degli stimoli (bi- o tri-dimensionali), grafico a dispersione delle disparità rispetto alle distanze.

Considerazioni sui dati dello Scaling multidimensionale

Dati. Se si dispone di dati di dissimilarità, tutte le dissimilarità dovrebbero essere quantitative e dovrebbero essere misurate in base alla stessa metrica. Se i dati sono multivariati, le variabili possono essere quantitative, binarie o dati di conteggio. Lo scaling delle variabili è una questione importante poiché le differenze possono influenzare la soluzione. Se le variabili hanno differenze significative (ad esempio, una variabile è misurata in dollari e l'altra è misurata in anni), è consigliabile standardizzarle. Questa operazione può essere eseguita automaticamente dalla procedura Scaling multidimensionale.

Ipotesi. La procedura Scaling multidimensionale è relativamente libera da ipotesi di distribuzione. Assicurarsi di selezionare il livello di misurazione appropriato (ordinale, intervallo o rapporto) nella finestra di dialogo Scaling multidimensionale per essere sicuri che i risultati vengano calcolati correttamente.

Procedure correlate. Se l'obiettivo è la riduzione dei dati, si può considerare un metodo alternativo quale l'analisi fattoriale, in particolare se le variabili sono di tipo quantitativo. Se si desidera identificare gruppi di casi simili, considerare la possibilità di integrare l'analisi di scaling multidimensionale con un'analisi gerarchica o cluster *k*-medie.

Per ottenere un'analisi Scaling multidimensionale

1. Dai menu, scegliere:

   Analizza > **Scala** > **Scaling multidimensionale...**

2. Selezionare almeno quattro variabili numeriche per l'analisi.
3. Nel gruppo Distanze selezionare **I dati sono distanze** o **Crea le distanze dai dati**.
4. Se si seleziona **Crea le distanze dai dati**, è anche possibile selezionare una variabile di raggruppamento per le singole matrici. La variabile di raggruppamento può essere di tipo numerico o stringa.

In alternativa, è possibile anche:

- Specificare una forma della matrice di distanza quando i dati sono distanze.
- Specificare la misura di distanza da utilizzare durante la creazione delle distanze dai dati.

Utilizzo della mappatura di prossimità per il ridimensionamento multidimensionale

Oltre ad ALSCAL, IBM SPSS Statistica offre **Proximity Mapping (PROXMAP)** come approccio correlato al ridimensionamento multidimensionale. ALSCAL e PROXMAP differiscono nel modo in cui viene calcolata l'adattamento del modello. ALSCAL adatta le distanze al quadrato alle dissimilarità al quadrato, mentre PROXMAP adatta le distanze direttamente alle dissimilarità. Di conseguenza, le due procedure possono fornire valori numerici diversi anche quando le configurazioni spaziali risultanti sembrano simili. Queste differenze possono risultare più evidenti quando si lavora con dati per i quali non esiste una soluzione quasi perfetta.

Mentre ALSCAL offre una scalatura multidimensionale di base con un output testuale limitato e informazioni diagnostiche minime, PROXMAP propone soluzioni moderne e complete con visualizzazioni relativamente accurate (tra cui diagrammi di Shepard, grafici di trasformazione e analisi dei residui), un output statistico potenziato (misure di stress multiple, valori R-quadrato e indici di adattamento dettagliati) e funzionalità avanzate come l'adattamento delle proprietà.

Consulta i seguenti link per iniziare a utilizzare Proximity Mapping:

- “Confronto tra ALSCAL e PROXMAP” a pagina 272
- Associazione di prossimità
-
-
-

Scaling multidimensionale: Forma dei dati

Se il dataset attivo rappresenta le distanze tra uno o due insiemi di oggetti, è necessario specificare la forma della matrice dei dati in modo da ottenere i risultati corretti.

Nota: non è possibile selezionare **Quadrata simmetrica** se la finestra di dialogo Modello specifica la condizionalità della riga.

Scaling multidimensionale: Crea misure dai dati

La procedura Scaling multidimensionale utilizza dati di dissimilarità per creare una soluzione di scaling. Se i dati disponibili sono dati multivariati (valori di variabili misurate), è necessario creare dati di dissimilarità in modo da calcolare una soluzione di scaling multidimensionale. È possibile specificare i dettagli della creazione delle misure di dissimilarità a partire dai dati disponibili.

Misura. Consente di specificare la misura di dissimilarità per l'analisi. Selezionare un'alternativa dal gruppo Misura corrispondente al tipo di dati desiderato e quindi selezionare una delle misure dall'elenco a discesa corrispondente a tale tipo di misura. Le alternative disponibili sono:

- **Intervallo**. Distanza euclidea, Distanza euclidea al quadrato, Chebychev, City-Block, Minkowski o Personalizzato.
- **Conteggi**. Misura chi-quadrato e Misura phi-quadrato.
- **Binaria**. Distanza euclidea, Distanza euclidea al quadrato, Differenza di dimensione, Differenza di modello, Varianza o Lance e Williams.

Crea matrice delle distanze. Consente di scegliere l'unità di analisi. Le alternative sono Fra variabili o Fra casi.

Trasforma valori. In alcuni casi, ad esempio quando le variabili sono misurate su scale molto diverse, è possibile standardizzarne i valori prima di calcolare le prossimità (non applicabile ai dati binari). Selezionare un metodo di standardizzazione dall'elenco a discesa Standardizza. Se non è richiesta alcuna standardizzazione, selezionare **Nessuna**.

Scaling multidimensionale: Modello

La stima corretta di un modello di scaling multidimensionale dipende dagli aspetti dei dati e del modello stesso.

Livello di misurazione. Consente di specificare il livello dei dati. Le alternative sono Ordinale, Intervallo o Rapporto. Se le variabili sono ordinali, la selezione dell'opzione **Distingui le osservazioni correlate** richiede che vengano trattate come variabili continue, in modo che le correlazioni (valori uguali per casi differenti) siano risolte in modo ottimale.

Condizionalità. Consente di specificare quali confronti sono significativi. Le alternative sono Matrice, Riga o Non condizionale.

Dimensioni. Consente di specificare la dimensione della soluzione di scaling. Per ciascun numero nell'intervallo viene calcolata una soluzione. Specificare interi da 1 a 6. È consentito un minimo di 1 solo se si è selezionato **Distanza euclidea** come modello di scaling. Per una soluzione singola, specificare lo stesso valore minimo e massimo.

Modello di scaling. Consente di specificare le ipotesi in base alle quali viene eseguito lo scaling. Le alternative disponibili sono Distanza euclidea o Distanza euclidea per differenze singole (nota anche come INDSCAL). Nel modello Distanza euclidea per differenze singole, è possibile selezionare l'opzione **Accetta pesi del soggetto negativi**, se appropriata per i dati disponibili.

Scaling multidimensionale: Opzioni

È possibile impostare le opzioni per l'analisi scaling multidimensionale:

Visualizza. Consente di selezionare vari tipi di output. Le opzioni disponibili sono Grafici di gruppo, Grafici di soggetti singoli, Matrice dei dati e Informazioni sul modello e sulle opzioni.

Criteri. Consente di determinare il momento in cui interrompere l'iterazione. Per modificare i valori predefiniti, inserire i valori per **Convergenza s-stress**, **Valore minimo di s-stress** e **Massimo numero di iterazioni**.

Considera le distanze minori di: n come mancanti. Le distanze minori di tale valore sono escluse dall'analisi.

Funzioni aggiuntive del comando ALSCAL

Il linguaggio della sintassi dei comandi consente inoltre di:

- Utilizzare tre tipi di modelli aggiuntivi, noti come ASCAL, AINDS e GEMSCAL, per lo scaling multidimensionale.
- Eseguire trasformazioni polinomiali su dati di intervallo e di rapporto.
- Analizzare le similarità (piuttosto che le distanze) con dati ordinali.
- Analizzare i dati nominali.
- Salvare varie matrici di coordinate e pesi nei file e rileggerli per l'analisi.
- Vincolare l'unfolding multidimensionale.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Confronto tra ALSCAL e PROXMAP

Sia ALSCAL che PROXMAP eseguono un ridimensionamento multidimensionale (MDS) per rappresentare i dati di prossimità in un numero specificato di dimensioni. In PROXMAP, questi risultati vengono presentati utilizzando una terminologia aggiornata e grafici diagnostici che consentono un confronto diretto con i risultati di ALSCAL.

ALSCAL e PROXMAP stimano la disposizione spaziale degli oggetti sulla base dei dati di prossimità. PROXMAP offre un framework basato su finestre di dialogo che supporta gli stessi risultati concettuali di ALSCAL, utilizzando una terminologia aggiornata e grafici diagnostici consolidati.

Mappatura concettuale

La tabella seguente illustra la corrispondenza tra i termini chiave e i risultati di ALSCAL e la terminologia di PROXMAP. Il significato analitico dei risultati rimane invariato indipendentemente dalla procedura utilizzata.

Tabella 5. Mappatura terminologica tra ALSCAL e PROXMAP

Termine o risultato ALSCAL	**Equivalente a PROXMAP**
Configurazione stimolo derivata	Spazio degli oggetti comune
Grafico a dispersione di adattamento lineare Disparità Distanze	**Fit Source <Etichetta del produttore>** Prossimità trasformate Distanze
Grafico a dispersione di adattamento non lineare Osservazioni Distanze	**Grafico di Shepard[1]Fonte <Etichetta fonte>** Prossimità originali Distanze + Prossimità trasformate
Grafico a dispersione di trasformazione Osservazioni Disparità	**Fonte della trasformazione <Etichetta della fonte>** Prossimità originali Prossimità trasformate

Passaggio da ALSCAL a PROXMAP

Il seguente tutorial illustra come riprodurre un'analisi di scalatura multidimensionale (MDS) ALSCAL utilizzando PROXMAP e spiega come interpretare le somiglianze e le differenze nei risultati.

Osservazioni generali

Questo tutorial illustra come configurare PROXMAP per replicare un'analisi ALSCAL e come confrontare le configurazioni risultanti.

L'esempio si concentra su un unico set di dati di prossimità e utilizza impostazioni predefinite o equivalenti, in modo che le differenze tra ALSCAL e PROXMAP si limitino al ridimensionamento e all'orientamento, anziché comportare modifiche sostanziali nella soluzione.

[1] Il termine «Shepard Plot» si riferisce a Roger Shepard, uno dei fondatori del Multidimensional Scaling non metrico. Combina il grafico "Adattamento non lineare" di ALSCAL e il grafico "Trasformazione".

Dati utilizzati nell'analisi

L'analisi si basa su Dist_, un set di dati di prossimità che rappresenta le distanze di volo tra dodici città degli Stati Uniti. I dati vengono memorizzati sotto forma di matrice di prossimità simmetrica in forma triangolare inferiore.

Nella vista Dati, il set di dati rappresenta una matrice di prossimità simmetrica di tipo " 12x12 " memorizzata in formato triangolare inferiore e una variabile identificativa dell'oggetto di tipo nominale utilizzata per etichettare la configurazione.

Nella vista "Variabili", il set di dati è costituito da dodici variabili quantitative che contengono la matrice di prossimità e da una variabile nominale che definisce le etichette degli oggetti.

La variabile *CITY* è specificata come variabile di etichetta dell'oggetto.

Oltre alla matrice di prossimità, il set di dati include una variabile nominale di identificazione degli oggetti (*CITY*) che fornisce le etichette per gli oggetti presenti nella configurazione.

I valori di *CITY* vanno da 1 a 12. A questi numeri interi vengono assegnati dei valori, dove:

- 1 = ATLANTA
- 2 = BOSTON
- 3 = CINCINATI
- 4 = COLUMBUS
- 5 = DALLAS
- 6 = INDNPLIS
- 7 = LITTROCK
- 8 = LOSANGEL
- 9 = MEMPHIS
- 10 = ST. LOUIS
- 11 = SPOKANE
- 12 = TAMPA

PROXMAP utilizza queste etichette di valore per identificare gli oggetti nelle tabelle e nei risultati grafici, garantendo che vengano visualizzati i nomi degli oggetti anziché i codici numerici.

Configurazione dell'analisi PROXMAP

Vai alla **mappatura** > **della** > **prossimità in AnalyzeMapping**

Le variabili della matrice di prossimità vengono specificate come un'unica fonte di prossimità nella finestra di dialogo "Mappatura di prossimità". PROXMAP accetta dati di prossimità memorizzati sia sotto forma di matrice triangolare inferiore (diagonale inclusa) sia come matrice simmetrica completa, a condizione che la matrice sia specificata correttamente.

Scheda Dati (Impostazione della matrice di prossimità e delle etichette degli oggetti)

Definire la sorgente di prossimità:

1. Seleziona le variabili da *ATLANTA* a *TAMPA* nell'elenco Variabili e spostale nella **sezione** Variabili dell'origine.
2. Inserisci un nome nel campo "**Etichetta origine** " (ad esempio, "*Distanze di volo tra le città* ").
3. Fai clic su **"Aggiungi fonte"** per creare una fonte (ad esempio, *$SRC001*).
4. In "**Tipo di dati** ", seleziona **"Dati di prossimità"**.
5. Sposta la variabile *CITY* nelle **etichette degli oggetti**

Il numero attuale di oggetti = 12 corrisponde al numero di oggetti definiti dalla matrice di prossimità e conferma che nella configurazione saranno rappresentate dodici città.

Scheda Modello (Impostazione del modello PROXMAP)

L'analisi si basa su un'unica sorgente di prossimità ($SRC001), pertanto selezionare le seguenti opzioni:

1. Nel **modello a tre vie**, selezionare l'opzione **Identità**.
2. Nella sezione **"Condizionalità / Trasformazione"**, selezionare **"Senza condizioni"**
3. Fai clic su **«Definisci trasformazione»** e seleziona le opzioni **«Ordinale»** e «**Consenti di sciogliere i pareggi** ».

   Ciò indica che le prossimità vengono considerate come dissimilarità ordinali. L'ordine di classificazione delle distanze viene mantenuto e i valori uguali nei dati rimangono a pari merito nella trasformazione monotona.

   Se necessario, è possibile modificare il tipo di trasformazione (ad esempio, lineare, esponenziale, spline e così via).

Il numero attuale di dimensioni = 2 indica che verrà stimata una configurazione bidimensionale.

Questa configurazione corrisponde alle specifiche utilizzate per il confronto con ALSCAL.

Scheda Criteri (Impostazione della dimensionalità, della convergenza e della standardizzazione)

1. **Dimensionalità** : il **numero di dimensioni** è impostato su *2*

   Ciò indica che verrà stimata una configurazione bidimensionale.
2. **Convergenza dell'algoritmo** : i criteri di convergenza regolano il processo di ottimizzazione iterativa. Sono specificate le seguenti impostazioni:

   - **Funzione minima:** 0.000001
   - **Differenze nelle funzionalità:** 0.000000000001
   - **Differenze di configurazione:** 0.000001
   - **Numero massimo di iterazioni:** 1000

   Questi valori determinano quando l'algoritmo iterativo si interrompe. La convergenza si raggiunge quando il miglioramento della funzione di perdita e le variazioni nella configurazione scendono al di sotto delle tolleranze specificate, oppure quando viene raggiunto il numero massimo di iterazioni.
3. **Standardizzazione** : selezionare "**Somma dei quadrati = N** ". Questa opzione è particolarmente utile per la standardizzazione dei dati multivariati.
4. **Dimensione** : seleziona "**Per dimensione** ". Con questa opzione è possibile visualizzare le singole dimensioni. Selezionare la casella di controllo sotto **"Riflessione"** per la dimensione 2.

   I riflessi influenzano solo l'orientamento (il segno) degli assi e non modificano le distanze né l'adattamento del modello. Vengono solitamente utilizzati per allineare la soluzione a una configurazione di riferimento (ad esempio, quando si effettua un confronto con una soluzione ALSCAL).

Scheda "Output": sottoscheda "Modello" (Richiesta di tabelle e grafici)

1. Nella sezione Tabelle, selezionare **Coordinate** in modo che nell'output venga generata una tabella contenente le coordinate dell'oggetto nello spazio comune.

   Le opzioni **"Direzioni variabili"** e **"Pesi spaziali"** non sono selezionate, poiché questa analisi si basa su un'unica sorgente di prossimità e non coinvolge sorgenti aggiuntive né variabili multivariate.
2. Nella sezione "Grafici", all'interno della sezione "Spazio comune", seleziona **"Configurazione"**.

   Viene generato un grafico della configurazione dell'oggetto per la sorgente $SRC001 (ad esempio, le distanze di volo tra le città), con le etichette indicate nella variabile specificata nella scheda **Dati** (ad esempio, *CITY*).
3. **Azioni aggiuntive** : dall'elenco "**Fonti** ", spostare " *$SRC001* " nelle sezioni " **Trasformazione** ", "**Adattamento** " e "**Shepard** ".

Vengono generati i seguenti grafici:

- Un grafico che mostra **le distanze trasformate** sull'asse verticale rispetto alle distanze originali sull'asse orizzontale.
- Un grafico a dispersione che mostra le **distanze stimate** nella configurazione in funzione delle **prossimità trasformate**.
- Un grafico a dispersione diagnostico che mostra le distanze adattate, insieme alle prossimità trasformate rispetto a quelle originali, per visualizzare la bontà dell'adattamento nell'analisi multidimensionale non metrica (mappatura delle prossimità).

Si richiede una soluzione bidimensionale, in linea con la dimensionalità utilizzata nell'analisi ALSCAL.

Sintassi dei comandi

Inoltre, si veda la seguente sintassi di comando utilizzata per l'analisi tramite ALSCAL e PROXMAP.

Sintassi dei comandi utilizzata per l'analisi 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.
```

Sintassi dei comandi utilizzata per l'analisi 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)).
```

Grafici di output e diagnostici

L'output comprende una tabella delle coordinate degli oggetti nello spazio comune degli oggetti e un grafico di configurazione che mostra la rappresentazione spaziale delle città.

Si richiedono ulteriori grafici diagnostici, tra cui grafici di trasformazione, grafici di adattamento e grafici di Shepard. Questi grafici mostrano le relazioni tra le prossimità originali, le prossimità trasformate e le distanze adattate, e forniscono informazioni visive sulla qualità dell'adattamento della soluzione.

La terminologia di PROXMAP differisce da quella di ALSCAL. Ad esempio, gli stimoli ALSCAL vengono definiti «oggetti» in PROXMAP, mentre le disparità ALSCAL vengono definite «prossimità trasformate». Il grafico di Shepard in PROXMAP riunisce le informazioni riportate nei grafici di adattamento non lineare e di trasformazione di ALSCAL.

Risultati e interpretazione

Quando PROXMAP viene eseguito utilizzando impostazioni simili a quelle di ALSCAL, le configurazioni risultanti sono molto simili. Le differenze tra le soluzioni si limitano a un fattore di scala generale e a eventuali riflessioni degli assi.

I valori delle coordinate possono differire numericamente, ma rappresentano la stessa configurazione di base, fatta salva la scala. PROXMAP riporta le coordinate in base alla scala originale delle vicinanze.

Un confronto visivo dei grafici di configurazione mostra che le posizioni relative degli oggetti vengono mantenute, a conferma del fatto che PROXMAP riproduce correttamente la soluzione di ALSCAL.

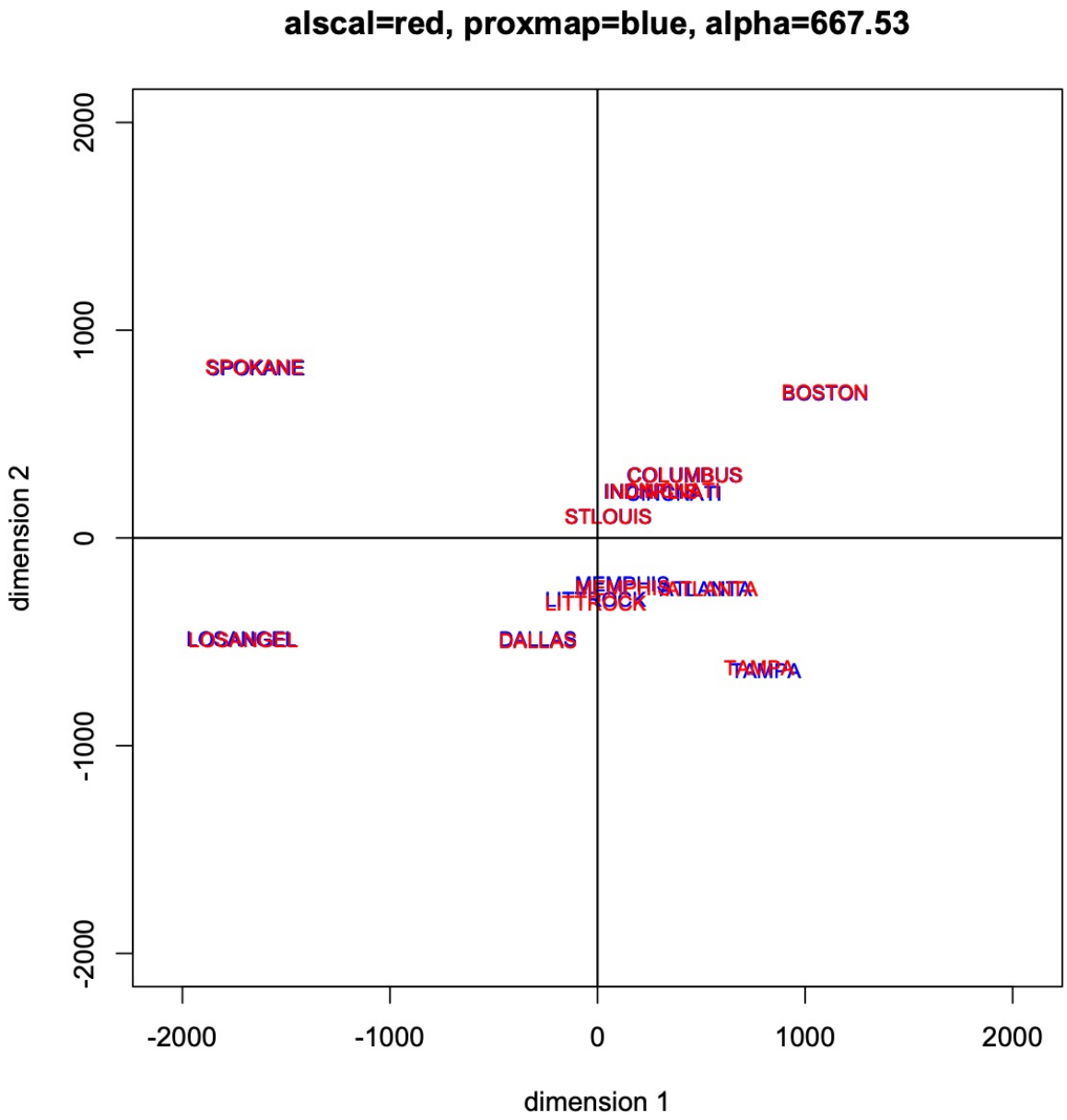


Figura 3. Configurazione dello stimolo ALSCAL Coordinate PROXMAP Spazio comune

Punti chiave per gli utenti di ALSCAL

- I dati di prossimità esistenti per ALSCAL possono essere riutilizzati in PROXMAP, compresi quelli memorizzati sotto forma di matrice triangolare inferiore.
- È necessaria un'ulteriore variabile di etichetta degli oggetti per identificare gli oggetti nell'output e nei grafici di PROXMAP.
- Quando si confrontano i risultati di PROXMAP con quelli di ALSCAL, possono verificarsi differenze nell'orientamento delle coordinate. I riflessi influiscono solo sull'orientamento degli assi e non modificano le distanze né la bontà dell'adattamento.

- PROXMAP utilizza una terminologia aggiornata e moderni metodi di scalatura multidimensionale, pur mantenendo la continuità concettuale con ALSCAL.

Statistiche dei rapporti

La procedura Statistica dei rapporti offre un elenco completo di statistiche di riepilogo per la descrizione del rapporto tra due variabili di scala.

È possibile ordinare l'output in base ai valori di una variabile di raggruppamento in ordine crescente o decrescente. È possibile eliminare il report sulla statistica dei rapporti dall'output e salvare i risultati in un file esterno.

Esempio
Esiste un buon grado di uniformità nel rapporto tra prezzo di valutazione e prezzo di vendita delle case in ognuno dei cinque paesi? Dall'output, si apprende che la distribuzione dei rapporti varia notevolmente da paese a paese.

Statistiche
Mediana, media, media ponderata, intervalli di confidenza, coefficiente di dispersione (COD), coefficiente di variazione (COV), coefficiente di variazione medio - centrato di variazione, coefficiente di variazione medio - correlato (PRB), differenziale di prezzo (PRD), deviazione standard, deviazione standard media assoluta (AAD), range, valori minimi e massimi, e l'indice di concentrazione calcolato per una gamma o percentuale specificata all'interno del rapporto mediano.

Considerazioni sui dati

Dati
Utilizzare codici numerici o stringhe per codificare le variabili di raggruppamento (misure di livello nominale o ordinale).

Ipotesi
Le variabili che definiscono il numeratore e il denominatore del rapporto dovrebbero essere variabili di scala con valori positivi.

Ottenere Rapporto Statistiche

1. Dai menu, scegliere:

   Analizza > **Statistiche descrittive** > **Rapporto...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una variabile numeratore.
3. Selezionare una variabile denominatore.

Oppure:

- Selezionare una variabile di raggruppamento e specificare l'ordinamento dei gruppi nei risultati.
- Scegliere se visualizzare o meno i risultati nel Visualizzatore.
- Scegliere di aggiungere una parola chiave **N** al sottocomando **PRINT** . L'impostazione aggiunge le dimensioni del campione alla tabella delle statistiche nell'output.
- Decidere se salvare o meno i risultati in un file esterno per futuri utilizzi. In caso affermativo, specificare il nome del file in cui verranno salvati.

Statistiche dei rapporti: Statistiche

Tendenza centrale
Misure di tendenza centrale sono statistiche che descrivono la distribuzione dei rapporti.

Mediana
Il valore tale che il numero di rapporti inferiore a questo valore e il numero di rapporti maggiore di questo valore siano uguali.

Media
Il risultato di sommare i rapporti e suddividere il risultato per il numero totale di rapporti.

Media pesata
Il risultato di dividere la media del numeratore per la media del denominatore. La media pesata è anche la media dei rapporti pesati dal denominatore.

Intervalli di confidenza per PRB e misure di tendenza centrale
Visualizza gli intervalli di confidenza per PRB, la media, la mediana e la media ponderata (se richiesto). Specificare un valore maggiore o uguale a 0 e minore di 100 come intervallo di confidenza. L'impostazione predefinita è 95%.

Dispersione
Queste statistiche misurano la quantità di variazione, o diffusione, nei valori osservati.

AAD
(Deviazione assoluta media) È ottenuta sommando le deviazioni assolute dei rapporti dalla mediana e dividendo il risultato per il numero totale di rapporti.

COD
(Coefficiente di dispersione) È il risultato dell'espressione della deviazione assoluta media come una percentuale della mediana.

CV
CV indica il coefficiente di variazione. È il rapporto tra la deviazione standard e la media. I valori sono espressi in percentuale.

PRB
Il bias legato al prezzo. Un indice del fatto che la valutazione ai rapporti di prezzo sia sistematicamente superiore o inferiore per le proprietà ad alto prezzo. Regala differenze percentuali nei rapporti di valutazione dal rapporto mediano sulla base 2 logaritmi di misure proxy di valore calcolate come medie dei prezzi di vendita e i rapporti dei valori valutati al rapporto mediano. Dà la variazione percentuale dei rapporti di valutazione per una variazione del valore del 100.

PRD
(Differenziale di prezzo) Anche noto come indice di regressività, è il risultato della divisione della media per la media pesata.

Deviazione standard
La deviazione standard viene ottenuta sommando le deviazioni quadrate dei rapporti dalla media, dividendo il risultato per il numero totale di rapporti meno uno ed estraendo la radice quadrata positiva.

Intervallo
La gamma è il risultato di sottrarre il rapporto minimo dal rapporto massimo.

Minimo
Il minimo è il rapporto più piccolo.

Massimo
Il massimo è il rapporto più grande.

Indice di concentrazione
Consente di misurare la percentuale di rapporti che cade in un intervallo. È possibile calcolarlo in due modi diversi:

Tra proporzioni
In questo caso l'intervallo viene definito in modo esplicito specificando i valori alti e bassi dell'intervallo. Immettere i valori per le proporzioni alte e basse, quindi fare clic su **Aggiungi** per ottenere un intervallo.

Entro percentuale della mediana
In questo caso l'intervallo viene definito in modo implicito specificando la percentuale della mediana. Digitare un valore compreso tra 0 e 100 e fare clic su **Aggiungi**. L'estremità inferiore dell'intervallo è uguale a (1 – 0.01 × valore) × mediana e l'estremità superiore è uguale a (1 + 0.01 × valore) × mediana.

Grafico P-P

La procedura di plotti P-P produce complotti di probabilità di una o più variabili di sequenza o di serie temporali. Le variabili possono essere standardizzate, differenziate e trasformate prima del complotto.

Le distribuzioni di test disponibili comprendono beta, chi-quadrato, esponenziale, gamma, semi-normale, Laplace, logistica, lognormale, normale, pareto, t di Student, Weibull e uniforme. A seconda della distribuzione selezionata, è possibile specificare i gradi di libertà e altri parametri.

- È possibile ottenere trasformare grafici delle probabilità per i valori trasformati. Le opzioni di trasformazione comprendono logaritmo naturale, valori standardizzati, differenza e differenza stagionale.
- È possibile specificare il metodo per calcolare le distribuzioni previste e per risolvere le "correlazioni" (più osservazioni con lo stesso valore).

Distribuzione del test
Specificare un tipo di distribuzione per i tuoi dati. L'elenco a discesa fornisce le seguenti opzioni:

Beta
Distribuzione beta. I parametri shape1 e shape2 *a* e *b* devono essere positivi. Se non sono specificati, `DISTRIBUTION` li stima dalla deviazione standard media e campione. Tutte le osservazioni devono essere comprese nella gamma 0 - 1, inclusive.

Chi-quadrato
Distribuzione Chi - quadrato. È necessario specificare i gradi di libertà (*df)*. Le osservazioni negative non sono consentite.

Esponenziale
Distribuzione esponenziale. Il parametro di scala *a* deve essere positivo. Se il parametro non è specificato, `DISTRIBUTION` lo stima dalla media del campione. Le osservazioni negative non sono consentite.

Gamma
Distribuzione gamma. I parametri di forma e scala *a* e *b* devono essere positivi. Se non sono specificati, `DISTRIBUTION` li stima dalla deviazione standard media e campione. Le osservazioni negative non sono consentite.

Semi-normale
Distribuzione mezza normale. Si ipotizza che i dati siano location - free o centralizzati. (Posizione parameter=0.) È possibile specificare il parametro di scala *a* o let `DISTRIBUTION` stimarlo utilizzando il metodo di massima verosimiglie.

Laplace
Laplace o doppia distribuzione esponenziale. `LAPLACE` prende una posizione e un parametro di scala (*a* e *b).* Il parametro di scala *(b*) deve essere positivo. Se i parametri non sono specificati, `DISTRIBUTION` li stima dalla deviazione standard media e campione.

Logistico
Distribuzione logistica. `LOGISTIC` prende una posizione e un parametro di scala (*a* e *b).* Il parametro di scala *(b*) deve essere positivo. Se i parametri non sono specificati, `DISTRIBUTION` li stima dalla deviazione standard media e campione.

Lognormale
Distribuzione Lognormale. I parametri di scala e di forma *a* e *b* devono essere positivi. Se non vengono specificati, `DISTRIBUTION` li stima dalla deviazione media e standard del logaritmo naturale dei dati di esempio. Le osservazioni negative non sono consentite.

Normale
Distribuzione normale. Il parametro di posizione *a* può essere qualsiasi valore numerico, mentre il parametro di scala *b* deve essere positivo. Se non sono specificati, DISTRIBUTION li stima dalla deviazione standard media e campione.

Pareto
Distribuzione Pareto. I parametri di soglia e di forma *a* e *b* devono essere positivi. Se non vengono specificati, DISTRIBUTION assume *a* uguale l'osservazione minima e le stime *b* con il metodo di massima verosimiglie. Le osservazioni negative non sono consentite.

T di Student
Distribuzione t *distribuzione.* È necessario specificare i gradi di libertà (*df).*

Uniforme
Distribuzione uniforme. UNIFORM prende un minimo e un parametro massimo *(a* e *b).* Il parametro *a* deve essere uguale o superiore a *b.* Se i parametri non sono specificati, DISTRIBUTION li assume dai dati di esempio.

Weibull
Distribuzione di Weibull. I parametri di scala e di forma *a* e *b* devono essere positivi. Se non vengono specificati, DISTRIBUTION li stima utilizzando il metodo meno quadrato. Le osservazioni negative non sono consentite.

Parametri di distribuzione
Fornisce la strategia di distribuzione e le opzioni dei parametri.

Stima dai dati
Quando selezionata, questa impostazione stima i parametri di distribuzione in base ai dati e al tipo di distribuzione selezionato.

Specificare
Quando selezionato, è possibile specificare i parametri di distribuzione per il tipo di distribuzione selezionato.

Nota: I parametri disponibili varieranno in base al tipo di distribuzione selezionato.

Trasforma
Le opzioni fornite impostano le impostazioni di trasformazione e periodicità.

Trasformazione logaritmica naturale
Trasforma i dati utilizzando il logaritmo naturale (base *e*) per rimuovere l'ampiezza varia.

Standardizza valori
Trasforma le variabili di serie o di serie temporali in un campione con una media di 0 e una deviazione standard di 1.

Differenza
Specifica il grado di differenziazione che viene utilizzato prima di tracciare per convertire una variabile non stazionaria in una variabile stazionaria con media e varianza costante. Inserire un valore appropriato nel campo.

Differenza stagionale
Se la variabile esibisce uno schema stagionale o periodico, è possibile utilizzare questa impostazione per differenza stagionale la variabile prima del complotto.

Nota: Questa impostazione è abilitata solo quando una variabile di serie o di serie temporali, con una periodicità definita, viene selezionata come una delle variabili quantitative.

Formula di stima della proporzione
Le opzioni fornite impostano la formula utilizzata per stimare le proporzioni.

Di Blom
Crea nuova variabile di graduatoria in base alle stime di proporzione che utilizza la formula `(r-3/8) / (w+1/4)`, dove `w` è la somma dei pesi dei casi e `r` è la classifica.

Rankit
Utilizza la formula `(r-1/2) / w`, dove `w` è il numero di osservazioni e `r` è la classifica, che varia da 1 a `w`.

Di Tukey
Utilizza la formula `(r-1/3) / (w+1/3)`, dove `r` è la classifica e `w` è la somma dei pesi dei casi.

Di Van der Waerden
La trasformazione di Van der Waerden, definita dalla formula `r/(w+1)`, dove `w` è la somma dei pesi dei casi e `r` è la classifica, che varia da 1 a `w`.

Rango assegnato alle correlazioni
Le opzioni fornite controllano il metodo per determinare come gestire i valori di tie. La tabella seguente mostra come i diversi metodi assegnano i ranghi ai valori legati.

Tabella 6. Metodi di classificazione e risultati

Valore	Media	Basso	Alto	Interrompi correlazioni arbitrariamente
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Ottenimento dei tracciati di probabilità P-P

1. Dai menu, scegliere:

   Analizzare > **Statistiche Descrittive** > **Pagine P-P ...**

   Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.
2. Selezionare una o più variabili numeriche e spostarle sul campo Variabili.
3. Selezionare una distribuzione del test.

Facoltativamente, è possibile selezionare le opzioni di trasformazione per ottenere complotti di probabilità per i valori trasformati e specificare un metodo per calcolare le distribuzioni previste.

Grafico Q-Q

La procedura di calcolo Q-Q produce complotti di probabilità per i valori trasformati. Le distribuzioni di test disponibili comprendono beta, chi-quadrato, esponenziale, gamma, semi-normale, Laplace, logistica, lognormale, normale, pareto, t di Student, Weibull e uniforme. A seconda della distribuzione selezionata, è possibile specificare i gradi di libertà e altri parametri.

- È possibile ottenere trasformare grafici delle probabilità per i valori trasformati. Le opzioni di trasformazione comprendono logaritmo naturale, valori standardizzati, differenza e differenza stagionale.
- È possibile specificare il metodo per calcolare le distribuzioni previste e per risolvere le "correlazioni" (più osservazioni con lo stesso valore).

Distribuzione del test
Specificare un tipo di distribuzione per i tuoi dati. L'elenco a discesa fornisce le seguenti opzioni:

Beta
Distribuzione beta. I parametri shape1 e shape2 *a* e *b* devono essere positivi. Se non sono specificati, `DISTRIBUTION` li stima dalla deviazione standard media e campione. Tutte le osservazioni devono essere comprese nella gamma 0 - 1, inclusive.

Chi-quadrato
Distribuzione Chi - quadrato. È necessario specificare i gradi di libertà (*df)*. Le osservazioni negative non sono consentite.

Esponenziale
Distribuzione esponenziale. Il parametro di scala *a* deve essere positivo. Se il parametro non è specificato, DISTRIBUTION lo stima dalla media del campione. Le osservazioni negative non sono consentite.

Gamma
Distribuzione gamma. I parametri di forma e scala *a* e *b* devono essere positivi. Se non sono specificati, DISTRIBUTION li stima dalla deviazione standard media e campione. Le osservazioni negative non sono consentite.

Semi-normale
Distribuzione mezza normale. Si ipotizza che i dati siano location - free o centralizzati. (Posizione parameter=0.) È possibile specificare il parametro di scala *a* o let DISTRIBUTION stimarlo utilizzando il metodo di massima verosimiglie.

Laplace
Laplace o doppia distribuzione esponenziale. LAPLACE prende una posizione e un parametro di scala (*a* e *b).* Il parametro di scala *(b*) deve essere positivo. Se i parametri non sono specificati, DISTRIBUTION li stima dalla deviazione standard media e campione.

Logistico
Distribuzione logistica. LOGISTIC prende una posizione e un parametro di scala (*a* e *b).* Il parametro di scala *(b*) deve essere positivo. Se i parametri non sono specificati, DISTRIBUTION li stima dalla deviazione standard media e campione.

Lognormale
Distribuzione Lognormale. I parametri di scala e di forma *a* e *b* devono essere positivi. Se non vengono specificati, DISTRIBUTION li stima dalla deviazione media e standard del logaritmo naturale dei dati di esempio. Le osservazioni negative non sono consentite.

Normale
Distribuzione normale. Il parametro di posizione *a* può essere qualsiasi valore numerico, mentre il parametro di scala *b* deve essere positivo. Se non sono specificati, DISTRIBUTION li stima dalla deviazione standard media e campione.

Pareto
Distribuzione Pareto. I parametri di soglia e di forma *a* e *b* devono essere positivi. Se non vengono specificati, DISTRIBUTION assume *a* uguale l'osservazione minima e le stime *b* con il metodo di massima verosimiglie. Le osservazioni negative non sono consentite.

T di Student
Distribuzione t *distribuzione.* È necessario specificare i gradi di libertà (*df).*

Uniforme
Distribuzione uniforme. UNIFORM prende un minimo e un parametro massimo *(a* e *b).* Il parametro *a* deve essere uguale o superiore a *b.* Se i parametri non sono specificati, DISTRIBUTION li assume dai dati di esempio.

Weibull
Distribuzione di Weibull. I parametri di scala e di forma *a* e *b* devono essere positivi. Se non vengono specificati, DISTRIBUTION li stima utilizzando il metodo meno quadrato. Le osservazioni negative non sono consentite.

Parametri di distribuzione
Fornisce la strategia di distribuzione e le opzioni dei parametri.

Stima dai dati
Quando selezionata, questa impostazione stima i parametri di distribuzione in base ai dati e al tipo di distribuzione selezionato.

Specificare
Quando selezionato, è possibile specificare i parametri di distribuzione per il tipo di distribuzione selezionato.

Nota: I parametri disponibili varieranno in base al tipo di distribuzione selezionato.

Trasforma
Le opzioni fornite impostano le impostazioni di trasformazione e periodicità.

Trasformazione logaritmica naturale
Trasforma i dati utilizzando il logaritmo naturale (base *e*) per rimuovere l'ampiezza varia.

Standardizza valori
Trasforma le variabili di serie o di serie temporali in un campione con una media di 0 e una deviazione standard di 1.

Differenza
Specifica il grado di differenziazione che viene utilizzato prima di tracciare per convertire una variabile non stazionaria in una variabile stazionaria con media e varianza costante. Inserire un valore appropriato nel campo.

Differenza stagionale
Se la variabile esibisce uno schema stagionale o periodico, è possibile utilizzare questa impostazione per differenza stagionale la variabile prima del complotto.

Nota: Questa impostazione è abilitata solo quando una variabile di serie o di serie temporali, con una periodicità definita, viene selezionata come una delle variabili quantitative.

Formula di stima della proporzione
Le opzioni fornite impostano la formula utilizzata per stimare le proporzioni.

Di Blom
Crea nuova variabile di graduatoria in base alle stime di proporzione che utilizza la formula `(r-3/8) / (w+1/4)`, dove `w` è la somma dei pesi dei casi e `r` è la classifica.

Rankit
Utilizza la formula `(r-1/2) / w`, dove `w` è il numero di osservazioni e `r` è la classifica, che varia da 1 a `w`.

Di Tukey
Utilizza la formula `(r-1/3) / (w+1/3)`, dove `r` è la classifica e `w` è la somma dei pesi dei casi.

Di Van der Waerden
La trasformazione di Van der Waerden, definita dalla formula `r/(w+1)`, dove `w` è la somma dei pesi dei casi e `r` è la classifica, che varia da 1 a `w`.

Rango assegnato alle correlazioni
Le opzioni fornite controllano il metodo per determinare come gestire i valori di tie. La tabella seguente mostra come i diversi metodi assegnano i ranghi ai valori legati.

Tabella 7. Metodi di classificazione e risultati

Valore	Media	Basso	Alto	Interrompi correlazioni arbitrariamente
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Ottenimento dei tracciati di probabilità Q-Q

1. Dai menu, scegliere:

   Analizzare > **Statistiche Descrittive** > **Q-Q grafici ...**

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una o più variabili numeriche e spostarle sul campo Variabili.
3. Selezionare una distribuzione del test.

Facoltativamente, è possibile selezionare le opzioni di trasformazione per ottenere complotti di probabilità per i valori trasformati e specificare un metodo per calcolare le distribuzioni previste.

Analisi ROC

L'analisi della caratteristica di funzionamento del ricevitore (ROC) è un modo utile per valutare l'accuratezza delle previsioni del modello tracciando la sensibilità versus (1-specificity) di un test di classificazione (in quanto la soglia varia su un'intera gamma di risultati dei test diagnostici). L'intera area sotto una data curva ROC, o AUC, formula una statistica importante che rappresenta la probabilità che la previsione sia nell'ordine corretto quando viene osservata una variabile di test (per un soggetto selezionato casualmente dal gruppo di casi e l'altro selezionato casualmente dal gruppo di controllo). L'analisi ROC supporta l'inferenza relativamente a un singolo AUC, curve richiamo precisione (PR) e fornisce opzioni per confrontare due curve ROC generate da gruppi indipendenti o soggetti accoppiati.

La vecchia procedura ROC Curve supporta l'inferenza statistica su una singola curva ROC. Questo può essere recuperato anche con la nuova procedura di analisi ROC. Inoltre, la nuova procedura di analisi ROC può confrontare due curve ROC generate da gruppi indipendenti o da soggetti accoppiati.

Le curve PR curve di precisione versus richiamo, tendono ad essere più informative quando i campioni di dati osservati sono altamente sfornati e forniscono un'alternativa alle curve ROC per i dati con un grande skew nella distribuzione di classe.

Esempio
E'nell'interesse di una banca classificare correttamente i clienti in quei clienti che faranno e non avranno default sui loro prestiti, quindi si sviluppano modelli speciali per prendere queste decisioni. L'analisi ROC può essere utilizzata per valutare e valutare l'accuratezza delle previsioni del modello.

Statistiche
AUC, gruppo negativo, valori mancanti, classificazione positiva, valore cutoff, forza di convinzione, intervallo di confidenza asintotica a due lati, distribuzione, errore standard, progettazione di gruppo indipendente, disegno a campione, assunzione non parametrica, assunzione di distribuzione esponenziale bi - negativa, midpoint, punto di taglio, curva PR, interpolazione stepwise, significato asintotico (2-tail), Sensibilità e (1-Specicity), Precisione e Recall.

Metodi
Si confronta le aree sotto due curve ROC, che vengono generate da gruppi indipendenti o da soggetti accoppiati. Confrontare due curve ROC può fornire maggiori informazioni nell'accuratezza derivate da due approcci diagnostici comparativi.

Considerazioni sui dati di analisi ROC

Dati
Le curve PR tracciano Precision versus Recall, e tendono ad essere più informative quando i campioni di dati osservati sono altamente sbruffati. Una semplice interpolazione lineare può generare erroneamente una stima troppo ottimistica di una curva PR.

Ipotesi
La previsione sarà nell'ordine corretto quando viene osservata una variabile di prova per un soggetto che viene selezionato casualmente dal gruppo di casi e l'altro viene selezionato casualmente dal gruppo di controllo. Ogni gruppo definito conterrà almeno una osservazione valida. Solo una singola variabile di raggruppamento viene utilizzata per una singola procedura.

Ottenimento di un'analisi ROC

1. Dai menu, scegliere:

Analizzare > **Classifica** > **Analisi ROC**

Nota: I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Seleziona uno o più **Prova variabile** variabili di probabilità del test.
3. Selezionane uno **Variabile di stato** .
4. Identifica il *positivo* **Valore della variabile di stato** .
5. Opzionalmente selezionare l'opzione **Pandate - sample design** oppure selezionare una singola variabile di raggruppamento (non è possibile selezionare entrambe le opzioni).
   - Utilizzare l'impostazione **Pandate - sample design** per richiedere la progettazione a campione accoppiato per la variabile di prova (s). Il design a campione abbinato mette a confronto due curve ROC in uno scenario a campione accoppiato quando vengono misurati più valori di test sugli stessi soggetti associati a una variabile di stato.

     Nota: Quando si seleziona **Pandate - sample design** , le opzioni **Raggruppamento Variabile** e **Distribuzione Assunzione** (nelle opzioni di dialogo Opzioni) sono disabilitate.
   - Quando viene selezionata una variabile di raggruppamento numerica, è possibile fare clic su **Definisci gruppi ...** per richiedere il design del gruppo indipendente per le variabili di prova, e per specificare due valori, un midpoint o un punto di taglio.
6. Opzionalmente, clicca su Opzioni per definire la classificazione, la direzione del test, i parametri di errore standard e le impostazioni dei valori mancanti.
7. Opzionalmente, clicca su Visualizza per definire le impostazioni di tracciamento e stampa (che includono ROC Curve, Precision - Recall Curve e impostazioni di qualità del modello).
8. Fare clic su **OK**.

Analisi ROC: Opzioni

È possibile selezionare le seguenti opzioni per l'analisi ROC:

Classificazione
. Consente di specificare se il valore di interruzione debba essere incluso o escluso dalla classificazione *positiva*. Questa impostazione non ha attualmente alcun effetto sull'output.

Direzione test
Fornisce opzioni per specificare quale direzione della variabile risultato del test indica aumentare la forza di convinzione che il soggetto è testato positivo.

Parametri per errore standard di area
. Consente di specificare il metodo di valutazione dell'errore standard dell'area sotto alla curva. I metodi disponibili sono esponenziali non parametrici e bi - negativi. L'impostazione predefinita **Nonparametric** fornisce le stime sotto l'assunzione non parametrica. L'impostazione **Bi - negativo esponenziale** fornisce le stime sotto l'assunzione di distribuzione esponenziale bi - negativa.

La sezione consente anche di specificare il livello di confidenza per l'intervallo di confidenza asintotico a due lati dell'AUC. L'intervallo disponibile è compreso tra 0.0% e 100.0% (il valore predefinito è 95%).

Nota: L'impostazione si applica solo al design del gruppo indipendente e non ha alcun effetto nel disegno a campione accoppiato.

Valori mancanti
Consente di specificare le modalità di gestione dei valori mancanti. Quando l'impostazione non viene selezionata, entrambi i valori mancanti dell'utente e i valori mancanti di sistema sono esclusi. Quando l'impostazione viene selezionata, i valori mancanti dell'utente come trattati come validi, sono esclusi i valori mancanti di sistema. I casi con valori mancanti di sistema, nella variabile di prova o nella variabile di stato, sono sempre esclusi dall'analisi.

Analisi ROC: Visualizzazione

È possibile specificare le seguenti impostazioni di visualizzazione per l'analisi ROC:

Grafico
Fornisce le opzioni per tracciare le curve ROC e Richiamo precisione.

Curva ROC
Quando selezionato, un grafico ROC Curve visualizza nell'output. Selezionare **Con diagonale di riferimento** per disegnare una linea di riferimento diagonale sul grafico ROC Curve.

Curva richiamo precisione
Quando selezionato, un grafico di Precision - Recall Curve visualizza nell'output. Precision - Rectutte le curve tendono ad essere più informative quando i campioni di dati osservati sono altamente sfornati e forniscono un'alternativa a ROC Curti per i dati con un grande skew nella distribuzione di classe. Il valore predefinito **Interpolato lungo l'impostazione dei veri positivi** rende la stepwise interpolazione lungo i veri positivi. L' **Interpolato lungo l'impostazione dei falsi positivi** rende la stepwise interpolazione lungo i falsi positivi.

Qualità generale del modello
L'impostazione controlla se viene creato o meno un grafico a barre per visualizzare il valore del limite inferiore dell'intervallo di confidenza dell'AUC stimata. Per impostazione predefinita, l'impostazione non è selezionata, che sopprime il grafico a barre.

Stampa
Fornisce le opzioni per la definizione dell'output per le statistiche corrispondenti.

Errore standard e intervallo di confidenza
L'impostazione controlla quale visualizzazione delle statistiche nella tabella "Area Sotto la curva". Quando l'impostazione non viene selezionata, solo i display AUC stimati. Quando l'impostazione viene selezionata, vengono visualizzate ulteriori statistiche, tra cui l'errore standard dell'AUC, il significato asintotico (2-tail) e i limiti dell'intervallo di confidenza Asintomtotic sotto l'ipotesi nulla.

Punti di coordinate della curva ROC
L'impostazione controlla i punti di coordinate del ROC Curva, insieme ai valori di cutoff. Quando l'impostazione non viene selezionata, viene soppressa l'emissione dei punti di coordinate. Quando l'impostazione viene selezionata, i valori delle coppie di Sensibilità e (1-Specificity) sono indicati con i valori di cutoff per ogni Curva ROC.

Indice di Youden

Visualizza il valore dell'indice di Youden per ogni valore di interruzione della curva ROC.

Punti di coordinate della curva Richiamo precisione
L'impostazione controlla i punti di coordinate della Precision - Recall Curve, insieme ai valori di cutoff. Quando l'impostazione non viene selezionata, viene soppressa l'emissione dei punti di coordinate. Quando l'impostazione viene selezionata, le coppie di valori di Precisione e Recensione sono date con i valori di cutoff per ogni curva di Precisione - Richiamo.

Metriche di valutazione del classificatore
L'impostazione controlla la visualizzazione della tabella Metriche di Valutazione dei Classifier nell'output. La tabella mostra come un modello di classificazione si adatta ai dati rispetto ad un assegnazione casuale e fornisce le seguenti informazioni:

- Le variabili di test specificate dall'utente
- Informazioni gruppo
- Gini Index (l'indice Gini è 2*AUC - 1, dove l'AUC è l'area sotto la curva ROC)
- Valori Max K-S e Cutoff

Analisi ROC: Definire i gruppi (stringa)

Per le variabili di raggruppamento delle stringhe, inserire una stringa per il Gruppo 1 e un altro valore per il Gruppo 2, come *sì* e *no*. I casi con altre stringhe sono esclusi dall'analisi.

Nota: I valori specificati devono esistere nella variabile, altrimenti un messaggio di errore visualizza per indicare che almeno uno dei gruppi è vuoto.

Analisi ROC: Definire i gruppi (numerici)

Per le variabili di raggruppamento numerico, definire i due gruppi per il test *t* specificando due valori, un midpoint o un punto di taglio.

Nota: I valori specificati devono esistere nella variabile, altrimenti un messaggio di errore visualizza per indicare che almeno uno dei gruppi è vuoto.

- **Usa i valori specificati.** Inserire un valore per il Gruppo 1 e un altro valore per il Gruppo 2. I casi con altri valori sono esclusi dall'analisi. Non è necessario specificare numeri interi (ad esempio, 6,25 e 12,5 sono validi).
- **Utilizza valore punto centrale.** Quando selezionate, i gruppi sono separati in valori < e ≥ midpoint.
- **Utilizza punto di divisione.**
  - **Punto di divisione.** Immettere un numero che suddivide i valori della variabile di raggruppamento in due insiemi Assegna i casi con valori minori al punto di divisione da un gruppo e i casi con valori maggiori o uguali al punto di divisione dall'altro gruppo.

Curva ROC

Questa procedura è utile per valutare la prestazione degli schemi di classificazione in cui i soggetti sono classificati in base a una variabile con due categorie.

Nota: Versione29.0.1 in poi, i menu e le finestre di dialogo per il comando ROC vengono rimossi e questo comando è disponibile solo come sintassi.

Esempio. Poiché è interesse della banca classificare in modo corretto i clienti adempienti o meno, è necessario elaborare metodi specifici per decidere a chi concedere i prestiti. Le curve ROC possono essere utilizzate per valutare il livello di attendibilità di questi metodi.

Statistiche. Area sotto alla curva ROC con intervallo di confidenza e coordinate della curva ROC. Grafici: Curva ROC.

Metodi. È possibile calcolare una stima dell'area sotto alla curva ROC in modo non parametrico o parametrico mediante un modello esponenziale binegativo.

Considerazioni sui dati della curva ROC

La vecchia procedura ROC Curve supporta l'inferenza statistica su una singola curva ROC. Questo può essere recuperato anche con la nuova procedura di analisi ROC. Inoltre, la nuova procedura di analisi ROC può confrontare due curve ROC generate da gruppi indipendenti o da soggetti accoppiati.

Dati. Le variabili del test sono quantitative. Le variabili del test sono spesso composte dalle probabilità dell'analisi discriminante o della regressione logistica oppure da punteggi su scala arbitraria, i quali rappresentano la "forza della convinzione" di uno stimatore di soggetti in una categoria piuttosto che in un'altra. La variabile di stato può essere di qualsiasi tipo e indica la vera categoria a cui appartiene un soggetto. Il valore della variabile di stato indica quale categoria debba essere considerata *positiva*.

Ipotesi. Si suppone che i numeri crescenti sulla scala dello stimatore rappresentino la convinzione crescente che il soggetto appartenga a una categoria, mentre i numeri decrescenti rappresentino la convinzione crescente che il soggetto appartenga all'altra categoria. L'utente deve scegliere qual è la direzione *positiva*. Si suppone inoltre che la categoria *vero* alla quale appartiene ciascun soggetto sia conosciuta.

Per creare una Curva ROC

1. Dai menù scegliere: **Analizzare** > **Classificare** > **Analisi ROC...**
2. Seleziona uno o più **Prova variabile** .
3. Selezionane uno **Variabile di stato** .

4. Individuare il valore *positivo* per la variabile di stato.
5. Clic **Schermo** , Selezionare **Curva ROC** E **Con linea di riferimento diagonale** , se necessario.
6. Clic **Continua**

Curva ROC: Opzioni

È possibile selezionare le seguenti opzioni per l'analisi ROC:

Classificazione
. Consente di specificare se il valore di interruzione debba essere incluso o escluso dalla classificazione *positiva*. Questa impostazione non ha attualmente alcun effetto sull'output.

Direzione test
. Consente di specificare la direzione della scala in rapporto alla categoria *positiva* .

Parametri per errore standard di area
. Consente di specificare il metodo di valutazione dell'errore standard dell'area sotto alla curva. I metodi disponibili sono esponenziale 'Non parametrico' ed 'Bi-negativo'. Consente inoltre di impostare il livello dell'intervallo di confidenza. L'intervallo disponibile è da 50,1% a 99,9%.

Valori mancanti
Consente di specificare le modalità di gestione dei valori mancanti.

Simulazione

I modelli predittivi, ad esempio la regressione lineare, richiedono un insieme di input noti per prevedere un risultato o un valore obiettivo. In molte applicazioni del mondo reale, tuttavia, i valori di input sono incerti. La simulazione consente di tenere conto dell'incertezza negli input per i modelli predittivi e valutare la possibilità di vari risultati del modello in presenza di tale incertezza. Ad esempio, si ha un modello di profitto che include il costo dei materiali come un input, ma vi è incertezza in tale costo a causa della volatilità del mercato. È possibile utilizzare la simulazione per modellare tale incertezza e determinare l'effetto che essa ha sul profitto.

La simulazione in IBM SPSS Statistica utilizza il metodo Monte Carlo. Gli input incerti vengono modellati con le distribuzioni della probabilità (ad esempio, la distribuzione triangolare) e i valori simulati per tali input vengono generati a partire da tali distribuzioni. Gli input i cui valori sono noti, vengono tenuti fissi sui valori noti. Il modello predittivo viene valutato utilizzando un valore simulato per ogni input incerto e i valori fissi per gli input noti per calcolare l'obiettivo (o gli obiettivi) del modello. Il processo viene ripetuto molte volte (in genere, decine di migliaia o centinaia di migliaia di volte), risultando in una distribuzione dei valori obiettivo che è possibile utilizzare per rispondere alle domande di natura probabilistica. Nel contesto di IBM SPSS Statistica , ogni ripetizione del processo genera un caso separato (record) di dati, composto dall'insieme di valori simulati per gli input incerti, dai valori degli input fissi e dall'obiettivo o dagli obiettivi previsti del modello.

È anche possibile simulare i dati in assenza di un modello predittivo specificando le distribuzioni di probabilità per le variabili che devono essere simulate. Ogni caso di dati generato consiste nell'insieme di valori simulati per le variabili specificate.

Per eseguire una simulazione, è necessario specificare i dettagli quali il modello predittivo, le distribuzioni della probabilità per gli input incerti, le correlazioni tra tali input e i valori per gli input fissi. Una volta specificati tutti i dettagli di una simulazione, è possibile eseguirla e, facoltativamente, salvare le specifiche in un file del **piano di simulazione**. È possibile condividere il piano di simulazione con altri utenti, che possono quindi eseguire la simulazione senza dover comprendere i dettagli del modo in cui è stata creata.

Sono disponibili due interfacce per utilizzare le simulazioni. Il Builder di simulazioni è un'interfaccia avanzata per gli utenti che progettano ed eseguono le simulazioni. Fornisce l'insieme completo di funzioni per la progettazione di una simulazione, per il salvataggio delle specifiche in un file del piano di simulazione, la specifica dell'output e l'esecuzione della simulazione. È possibile creare una simulazione in base a un file del modello IBM SPSS o a un insieme di equazioni personalizzate definite dall'utente nel Builder di simulazioni. Inoltre, è possibile caricare un piano di simulazione esistente nel Builder di simulazioni, modificare le impostazioni ed eseguire la simulazione, salvando facoltativamente il piano

aggiornato. Per gli utenti che hanno un piano di simulazione e desiderano principalmente eseguire la simulazione, è disponibile un'interfaccia più semplice. Essa consente di modificare le impostazioni che consentono di eseguire la simulazione in condizioni differenti, ma non offre le funzioni complete del Builder di simulazioni per la progettazione delle simulazioni.

Progettazione di una simulazione in base a un file del modello

1. Dai menu, scegliere:

   Analizzare > **Simulazione ...**
2. Fare clic su **Seleziona file di modello SPSS** e fare clic su **Continua**.
3. Aprire il file del modello.

   Il file del modello è un file XML contenente il modello PMML creato da IBM SPSS Statistica o IBM SPSS Modellatore. Per ulteriori informazioni, consultare l'argomento “Scheda Modello ” a pagina 291.
4. Nella scheda Simulazione (in Builder di simulazioni) specificare le distribuzioni di probabilità per gli input simulati e i valori per gli input fissi. Se il dataset attivo contiene i dati cronologici per gli input simulati, fare clic su **Adatta tutto** per determinare automaticamente la distribuzione che si adatta meglio ai dati per ciascuno di tali input e per determinare le correlazioni tra loro. Per ciascun input simulato che non viene adattato dai dati cronologici, è necessario specificare esplicitamente una distribuzione, selezionando un tipo di distribuzione e immettendo i parametri richiesti.
5. Fare clic su **Esegui** per eseguire la simulazione. Per impostazione predefinita, il piano di simulazione che specifica i dettagli della simulazione, viene salvato nell'ubicazione specificata nelle impostazioni di salvataggio.

Sono disponibili le seguenti opzioni:

- Modificare l'ubicazione del piano di simulazione salvato.
- Specificare correlazioni note tra gli input simulati.
- Calcolare automaticamente una tabella di contingenza di associazioni tra gli input categoriali e utilizzare queste associazioni quando i dati vengono generati per questi input.
- Specificare l'analisi di sensibilità per esaminare l'effetto che si ottiene modificando il valore di un input fisso o modificando un parametro di distribuzione di un input simulato.
- Specificare le opzioni avanzate, ad esempio l'impostazione del numero massimo di casi da generare o la richiesta del campionamento delle code.
- Personalizzare l'output.
- Salvare i dati simulati in un file di dati.

Progettazione di una simulazione in base a equazioni personalizzate

1. Dai menu, scegliere:

   Analizzare > **Simulazione ...**
2. Fare clic su **Digita le equazioni** e su **Continua**.
3. Fare clic su **Nuova equazione** nella scheda Modello (nel Builder di simulazioni) per definire ciascuna equazione nel modello predittivo.
4. Fare clic sulla scheda Simulazione e specificare le distribuzioni della probabilità per gli input simulati e i valori per gli input fissi. Se il dataset attivo contiene i dati cronologici per gli input simulati, fare clic su **Adatta tutto** per determinare automaticamente la distribuzione che si adatta meglio ai dati per ciascuno di tali input e per determinare le correlazioni tra loro. Per ciascun input simulato che non viene adattato dai dati cronologici, è necessario specificare esplicitamente una distribuzione, selezionando un tipo di distribuzione e immettendo i parametri richiesti.
5. Fare clic su **Esegui** per eseguire la simulazione. Per impostazione predefinita, il piano di simulazione che specifica i dettagli della simulazione, viene salvato nell'ubicazione specificata nelle impostazioni di salvataggio.

Sono disponibili le seguenti opzioni:

- Modificare l'ubicazione del piano di simulazione salvato.
- Specificare correlazioni note tra gli input simulati.
- Calcolare automaticamente una tabella di contingenza di associazioni tra gli input categoriali e utilizzare queste associazioni quando i dati vengono generati per questi input.
- Specificare l'analisi di sensibilità per esaminare l'effetto che si ottiene modificando il valore di un input fisso o modificando un parametro di distribuzione di un input simulato.
- Specificare le opzioni avanzate, ad esempio l'impostazione del numero massimo di casi da generare o la richiesta del campionamento delle code.
- Personalizzare l'output.
- Salvare i dati simulati in un file di dati.

Progettazione di una simulazione senza un modello predittivo

1. Dai menu, scegliere:

   Analizzare > **Simulazione ...**
2. Fare clic su **Crea dati simulati** e fare clic su **Continua**.
3. Nella scheda Modello (nel Builder di simulazioni), selezionare i campi che si desidera simulare. È possibile selezionare i campi dal dataset attivo oppure è possibile definire dei nuovi campi facendo clic su **Nuovo**.
4. Fare clic sulla scheda Simulazione e specificare le distribuzioni di probabilità per i campi che devono essere simulati. Se il dataset attivo contiene dei dati cronologici per uno qualsiasi di questi campi, fare clic su **Adatta tutto** per determinare automaticamente la distribuzione che si adatta meglio ai dati e per determinare le correlazioni tra i campi. Per i campi che non sono adatti ai dati cronologici, è necessario specificare esplicitamente una distribuzione selezionando un tipo di distribuzione e immettendo i parametri richiesti.
5. Fare clic su **Esegui** per eseguire la simulazione. Per impostazione predefinita, i dati simulati sono salvati nel nuovo dataset specificato nelle impostazioni di salvataggio. Inoltre, il piano di simulazione, che specifica i dettagli della simulazione, viene salvato nell'ubicazione specificata nelle impostazioni di salvataggio.

Sono disponibili le seguenti opzioni:

- Modificare l'ubicazione per i dati simulati oppure il piano di simulazione salvato.
- Specificare le correlazioni note tra i campi simulati.
- Calcolare automaticamente una tabella di contingenza di associazioni tra i campi categoriali e utilizzare queste associazioni quando i dati vengono generati per questi campi.
- Specificare l'analisi di sensibilità per esaminare l'effetto che si ottiene modificando un parametro di distribuzione per un campo simulato.
- Specificare le opzioni avanzate, ad esempio l'impostazione del numero di casi da generare.

Eseguire una simulazione da un piano di simulazione

Sono disponibili due opzioni per eseguire una simulazione da un piano di simulazione. È possibile utilizzare la finestra di dialogo Esegui simulazione, che è progettata principalmente per l'esecuzione da un piano di simulazione oppure è possibile utilizzare il Builder di simulazioni.

Per utilizzare la finestra di dialogo Esegui simulazione:

1. Dai menu, scegliere:

   Analizzare > **Simulazione ...**
2. Fare clic su **Apri un piano di simulazione esistente**.

3. Verificare che la casella di controllo **Apri in Builder di simulazioni** non sia selezionata e fare clic su **Continua**.
4. Aprire il piano di simulazione.
5. Fare clic su **Esegui** nella finestra di dialogo Esegui simulazione.

Per eseguire la simulazione dal Builder di simulazioni:

1. Dai menu, scegliere:

   Analizzare > **Simulazione ...**
2. Fare clic su **Apri un piano di simulazione esistente**.
3. Selezionare la casella di controllo **Apri in Builder di simulazioni** e fare clic su **Continua**.
4. Aprire il piano di simulazione.
5. Modificare le impostazioni desiderate nella scheda Simulazione.
6. Fare clic su **Esegui** per eseguire la simulazione.

Facoltativamente, è possibile eseguire quanto riportato di seguito.

- Impostare o modificare l'analisi di sensibilità per esaminare l'effetto che si ottiene modificando il valore di un input fisso o modificando un parametro di distribuzione di un input simulato.
- Riadattare le distribuzioni e le correlazioni per gli input simulati ai nuovi dati.
- Modificare la distribuzione per un input simulato.
- Personalizzare l'output.
- Salvare i dati simulati in un file di dati.

Builder di simulazioni

Il Builder di simulazioni offre l'insieme completo di funzioni per progettare ed eseguire le simulazioni. Consente di eseguire le attività generali riportate di seguito.

- Progettare ed eseguire una simulazione per un modello IBM SPSS definito in un file del modello PMML.
- Progettare ed eseguire una simulazione per un modello predittivo definito da un insieme di equazioni personalizzate specificate dall'utente.
- Progettare ed eseguire una simulazione che genera dati in assenza di un modello predittivo.
- Eseguire una simulazione in base a un piano di simulazione esistente, modificando facoltativamente le impostazioni del piano.

Scheda Modello

Per le simulazioni basate su un modello predittivo, la scheda Modello specifica l'origine del modello. Per le simulazioni che non includono un modello predittivo, la scheda Modello specifica i campi che devono essere simulati.

Seleziona file di modello SPSS. Questa opzione specifica che il modello predittivo è definito in un file del modello IBM SPSS. Un file del modello IBM SPSS è un file XML o un file archivio compresso (file .zip) contenente il modello PMML creato da IBM SPSS Statistica o IBM SPSS Modellatore. I modelli predittivi vengono creati dalle procedure, ad esempio Regressione lineare o Alberi decisionali all'interno di IBM SPSS Statistica , e possono essere esportati in un file del modello. È possibile utilizzare un file del modello differente facendo clic su **Sfoglia** e accedendo al file desiderato.

Modelli PMML supportati dalla simulazione

- Regressione lineare
- Modello lineare automatico
- Modello lineare generalizzato
- Modello lineare generalizzato misto

- Modello lineare generale
- Regressione logistica binaria
- Regressione logistica multinomiale
- Regressione multinomiale ordinale
- Regressione di Cox
- Albero
- Albero aumentato (C5)
- Discriminante
- Cluster TwoStep
- Cluster k-medie
- Rete neurale
- Insieme di regole (elenco decisionale)

Nota:

- I modelli PMML che hanno più campi obiettivo (variabili) o suddivisioni non sono supportati per l'uso nella simulazione.
- I valori di input stringa in modelli di regressione logistica binaria sono limitati a 8 byte nel modello. Se si stanno inserendo input stringa nel dataset attivo, assicurarsi che i valori nei dati non superino 8 byte in lunghezza. I valori di dati che superano 8 byte sono esclusi dalla dist ribuzione categoriale associata per l'input e sono visualizzati come non corrispondenti nella tabell a di output Categorie senza corrispondenza.

Digita le equazioni per il modello. Questa opzione specifica che il modello predittivo è composto da una o più equazioni personalizzate che verranno create dall'utente. Creare le equazioni facendo clic su **Nuova equazione**. Viene visualizzata la finestra di dialogo Editor di equazioni. È possibile modificare le equazioni esistenti, copiarle per utilizzarle come modelli per le nuove equazioni, registrarle ed eliminarle.

- Il Builder di simulazioni non supporta i sistemi di equazioni simultanee o le equazioni che sono non lineari nella variabile obiettivo.
- Le equazioni personalizzate vengono valutate nell'ordine in cui vengono specificate. Se l'equazione per un determinato obiettivo dipende da un altro obiettivo, l'altro obiettivo deve essere definito da un'equazione precedente.

  Ad esempio, dato l'insieme di tre equazioni fornito di seguito, l'equazione per *profitto* dipende dai valori di *ricavo* e *spese*, quindi le equazioni per *ricavo* e *spese* devono precedere l'equazione per *profitto*.

```
revenue = price*volume
expenses = fixed + volume*(unit_cost_materials + unit_cost_labor)
profit = revenue - expenses
```

Crea dati simulati senza un modello. Selezionare questa opzione per simulare i dati senza un modello predittivo. Specificare i campi che devono essere simulati selezionando i campi dal dataset attivo oppure facendo clic su **Nuovo** per definire i nuovi campi.

Editor di equazioni

L'Editor di equazioni consente di creare o di modificare un'equazione personalizzata per il modello predittivo.

- L'espressione per l'equazione può contenere dei campi dal dataset attivo o nuovi campi di input definiti dall'utente nell'Editor di equazioni.
- È possibile specificare le proprietà dell'obiettivo, quali il livello di misurazione, le etichette dei valori e se l'output viene generato o meno per l'obiettivo.
- È possibile utilizzare gli obiettivi delle equazioni definite in precedenza come input per l'equazione corrente, in modo da creare equazioni accoppiate.

- È possibile allegare un commento descrittivo all'equazione. I commenti vengono visualizzati insieme all'equazione nella scheda Modello.

1. Immettere il nome dell'obiettivo. Se lo si desidera, fare clic su **Modifica** nella casella di testo Obiettivo per aprire la finestra di dialogo Input definiti, che consente di modificare le proprietà predefinite dell'obiettivo.
2. Per creare un'espressione, incollare i componenti nel campo Espressione numerica o digitarli direttamente nel campo Espressione numerica.

- È possibile creare un'espressione personalizzata utilizzando i campi del dataset attivo oppure è possibile definire nuovi input facendo clic sul pulsante **Nuovo**. Viene visualizzata la finestra di dialogo Input definiti.
- È possibile incollare le funzioni selezionando un gruppo nell'elenco Gruppo di funzioni e facendo doppio clic sulla funzione nell'elenco Funzioni (o selezionare la funzione e fare clic sulla freccia accanto all'elenco Gruppo di funzioni). Immettere i parametri indicati dai punti interrogativi. Il gruppo di funzioni etichettato **Tutti** offre un elenco di tutte le funzioni disponibili. Una breve descrizione della funzione attualmente selezionata viene visualizzata in un'area riservata della finestra di dialogo.
- Le costanti stringa devono essere incluse tra virgolette.
- Se i valori contengono numeri decimali, è necessario utilizzare un punto (.) come indicatore decimale.

Nota: la simulazione non supporta equazioni personalizzate con le destinazioni stringa.

Input definiti

Nella finestra di dialogo Input definiti è possibile definire nuovi input e impostare le proprietà degli obiettivi.

- Se un input da utilizzare in un'equazione non esiste nel dataset attivo, è necessario definirlo prima che possa essere utilizzato nell'equazione.
- Se si stanno simulando dei dati senza un modello predittivo, è necessario definire tutti gli input simulati che non esistono nel dataset attivo.

Nome. Specificare il nome per un obiettivo o un input.

Obiettivo. È possibile specificare il livello di misurazione di un obiettivo. Il livello di misurazione predefinito è continuo. È anche possibile specificare se l'output verrà creato per questo obiettivo. Ad esempio, per un insieme di equazioni accoppiate è possibile che si sia interessati solo all'output prodotto dall'obiettivo per l'equazione finale, quindi si eliminerà l'output prodotto dagli altri obiettivi.

Input da simulare. Specifica che i valori dell'input verranno simulati in base a una distribuzione della probabilità specificata (la distribuzione della probabilità viene specificata nella scheda Simulazione). Il livello di misurazione determina l'insieme predefinito delle distribuzioni che vengono considerate durante la ricerca della distribuzione che meglio si adatta ai dati per l'input (facendo clic su **Adatta** o su **Adatta tutto** nella scheda Simulazione). Ad esempio, se il livello di misurazione è continuo, la distribuzione normale (appropriata per i dati continui) verrà considerata, ma la distribuzione binomiale non verrà considerata.

Nota: Selezionare un livello di misurazione di Stringa per gli input stringa. Gli input di stringa che devono essere simulati sono limitati alla distribuzione categoriale.

Input a valore fisso. Specifica che il valore dell'input è noto e verrà fissato sul valore noto. Gli input fissi possono essere di tipo numerico o stringa. Specificare un valore per l'input fisso. I valori stringa non devono essere racchiusi tra virgolette.

Etichette valori. È possibile specificare le etichette valori per gli obiettivi, gli input simulati e gli input fissi. Le etichette valore vengono utilizzate nei grafici di output e nelle tabelle.

Scheda Simulazione

La scheda Simulazione specifica tutte le proprietà della simulazione diverse dal modello predittivo. Nella scheda Simulazione è possibile effettuare le attività generali riportate di seguito.

- Specificare le distribuzioni di probabilità per gli input simulati e i valori per gli input fissi.
- Specificare le correlazioni tra gli input simulati. Per gli input categoriali, è possibile specificare che le associazioni che esistono tra questi input nel dataset attivo vengano utilizzate quando i dati sono generati per questi input.
- Specificare le opzioni avanzate, ad esempio, un campionamento delle code e i criteri di adattamento delle distribuzioni ai dati cronologici.
- Personalizzare l'output.
- Specificare se salvare il piano di simulazione e, facoltativamente, salvare i dati simulati.

Campi simulati

Per eseguire una simulazione, ogni campi di input deve essere specificato come fisso o simulato. Gli input simulati sono quelli i cui valori sono incerti e verranno generati a partire da una distribuzione della probabilità specificata. Quando sono disponibili i dati cronologici per gli input da simulare, le distribuzioni che si adattano meglio ai dati possono essere determinate automaticamente, insieme alle correlazioni tra tali input. Inoltre, è possibile specificare manualmente le distribuzioni o le correlazioni se i dati cronologici non sono disponibili o se sono necessarie delle distribuzioni o correlazioni specifiche.

Gli input fissi sono quelli i cui valori sono noti e rimangono costanti per ogni caso generato nella simulazione. Ad esempio, si dispone di un modello di regressione lineare per le vendite come una funzione di un numero di input tra cui il prezzo e si desidera tenere fisso il prezzo di mercato corrente. Quindi, si specificherà il prezzo come un input fisso.

Per le simulazioni basate su un modello predittivo, ogni predittore nel modello è un campo di input per la simulazione. Per le simulazioni che non includono un modello predittivo, i campi specificati nella scheda Modello sono gli input per la simulazione.

Adattamento automatico delle distribuzioni e calcolo delle correlazioni per gli input simulati. Se il dataset attivo contiene i dati cronologici per gli input che si desidera simulare, è possibile trovare automaticamente le distribuzioni che si adattano meglio ai dati per tali input e determinare le correlazioni tra loro. Le operazioni da eseguire vengono riportate di seguito.

1. Verificare che ciascuno degli input che si desidera simulare corrisponda al campo corretto nel dataset attivo. Gli input vengono elencati nella colonna Input e la colonna Adatta a visualizza il campo corrispondente nel dataset attivo. È possibile associare un input a un campo differente nel dataset attivo selezionando un elemento differente dall'elenco a discesa Adatta a.

   Il valore *-Nessuno-* nella colonna Adatta a indica che non è stato possibile associare automaticamente l'input a un campo nel dataset attivo. Per impostazione predefinita, gli input sono associati ai campi del dataset a livello di nome, misurazione e tipo (numerico o stringa). Se il dataset attivo non contiene i dati cronologici per l'input, specificare manualmente la distribuzione per l'input oppure specificare l'input come input fisso, come descritto di seguito.
2. Fare clic su **Adatta tutto**.

La distribuzione dell'adattamento più vicino e i parametri associati vengono visualizzati nella colonna Distribuzione, insieme a un grafico della distribuzione sovrapposto a un istogramma (o grafico grafico barre) dei dati cronologici. Le correlazioni tra gli input simulati vengono visualizzate nelle impostazioni di Correlazioni. È possibile esaminare i risultati dell'adattamento e personalizzare l'adattamento automatico della distribuzione per un particolare input selezionando la riga dell'input e facendo clic su **Dettagli adattamento**. Per ulteriori informazioni, consultare l'argomento “Dettagli adattamento ” a pagina 296.

È possibile eseguire l'adattamento automatico della distribuzione per un particolare input selezionando la riga dell'input e facendo clic su **Adatta**. Le correlazioni per tutti gli input simulati che corrispondono ai campi nel dataset attivo vengono anch'esse calcolate automaticamente.

Nota:

- I casi con valori mancanti per eventuale input simulato vengono esclusi dall'adattamento della distribuzione, dal calcolo delle correlazioni e dal calcolo della tabella di contingenza facoltativa (per input con una distribuzione categoriale). È possibile, facoltativamente, specificare se i valori mancanti definiti dall'utente di input con una distribuzione categoriale sono trattati come validi. Per impostazione

predefinita, vengono trattati come mancanti. Per ulteriori informazioni, consultare l'argomento “Opzioni avanzate ” a pagina 298.

- Per gli input continui e ordinali, se non viene trovato un adattamento accettabile per una qualsiasi delle distribuzioni testate, la distribuzione empirica viene suggerita come adattamento più vicino. Per gli input continui, la distribuzione empirica è la funzione di distribuzione cumulativa dei dati cronologici. Per gli input ordinali, la distribuzione empirica è la distribuzione categoriale dei dati cronologici.

Specifica manuale delle distribuzioni. È possibile specificare manualmente la distribuzione della probabilità per gli input simulati selezionando la distribuzione dall'elenco a discesa **Tipo** e immettendo i parametri di distribuzione nella griglia Parametri. Dopo avere immesso i parametri di una distribuzione, un grafico di esempio della distribuzione, in base ai parametri specificati, verrà visualizzato adiacente alla griglia Parametri. Di seguito vengono riportate alcune note relative a distribuzioni particolari.

- **Categoriale.** La distribuzione categoriale descrive un campo di input che ha un numero fisso di valori, a cui si fa riferimento come categorie. A ogni categoria è associata una probabilità, in modo che la somma delle probabilità in tutte le categorie sia uguale a uno. Per immettere una categoria, fare clic sulla colonna a sinistra nella griglia Parametri e specificare il valore categoria. Immettere la probabilità associata alla categoria nella colonna a destra.

  Nota: Gli input categoriali da un modello PMML hanno le categorie che sono determinate dal modello e non possono essere modificate.
- **Binomiale negativa - Errori.** Descrive la distribuzione del numero di errori in una sequenza di prove prima che venga osservato un numero specificato di esiti positivi. Il parametro *thresh* è il numero specificato di esiti positivi e il parametro *prob* è la probabilità di esito positivo in una determinata prova.
- **Binomiale negativa - Prove.** Descrive la distribuzione del numero di prove necessarie prima che venga osservato un numero specificato di esiti positivi. Il parametro *thresh* è il numero specificato di esiti positivi e il parametro *prob* è la probabilità di esito positivo in una determinata prova.
- **Intervallo.** Questa distribuzione consiste in una serie di intervalli con una probabilità assegnata ad ogni intervallo tale che la somma delle probabilità su tutti gli intervalli uguale a 1. I valori all'interno di un determinato intervallo sono tratti da una distribuzione uniforme definita in quell' intervallo. Gli intervalli sono specificati immettendo un valore minimo, un valore massimo e una probabilità associata.

  Ad esempio, si ritiene che il costo di una materia prima abbia il 40% di possibilità di rientrare nell'intervallo di $10 - $15 per unità e il 60% di possibilità di rientrare nell'intervallo di $15 - $20 per unità. L'utente modellerà il costo con una distribuzione dell'intervallo composta dai due intervalli [10 - 15] e [15 - 20], impostando la probabilità associata al primo intervallo su 0.4 e la probabilità associata al secondo intervallo su 0.6. Gli intervalli non devono essere continui e possono anche sovrapporsi. Ad esempio, è possibile specificare gli intervalli $10 - $15 e $20 - $25 o $10 - $15 e $13 - $16.
- **Weibull.** Il parametro *c* è un parametro di ubicazione facoltativo, che specifica dove è posizionata l'origine della distribuzione.

I parametri per le seguenti distribuzioni hanno un significato identico a quello nelle funzioni delle variabili casuali associate disponibili nella finestra di dialogo Calcola variabile: Bernoulli, beta, binomiale, esponenziale, gamma, lognormale, binomiale negativa (prove ed errori), normale, Poisson e uniforme.

Specifica di input fissi. Specificare un input fisso selezionando Fisso dall'elenco a discesa **Tipo** nella colonna Distribuzione e immettendo un valore fisso. Il valore può essere di tipo numerico o stringa a seconda che l'input sia di tipo numerico o stringa. I valori stringa non devono essere racchiusi tra virgolette.

Specifica dei limini nei valori simulati. La maggior parte delle distribuzioni supportano la specifica di limiti superiori e inferiori nei valori simulati. È possibile specificare un limite inferiore immettendo un valore nella casella di testo **Min** ed è possibile specificare un limite inferiore immettendo un valore nella casella di testo **Max**.

Blocco degli input. Bloccando un input selezionando la casella di controllo nella colonna con l'icona di blocco, si esclude l'input dall'adattamento automatico della distribuzione. Ciò è particolarmente utile quando si specifica manualmente una distribuzione o un valore fisso e si desidera verificare che non sia influenzata dall'adattamento automatico della distribuzione. Il blocco è utile anche se si desidera condividere il piano di simulazione con gli utenti che lo eseguiranno nella finestra di dialogo Esegui

simulazione e si desidera impedire che vengano modificati specifici input. A tal proposito, le specifiche per gli input bloccati non possono essere modificate nella finestra di dialogo Esegui simulazione.

Analisi di sensibilità. L'analisi di sensibilità consente di investigare l'effetto delle modifiche sistematiche in un input fisso o in un parametro di distribuzione per un input simulato generando un insieme indipendente di casi simulati - in modo efficace, una simulazione separata - per ogni valore specificato. Per specificare l'analisi di sensibilità, selezionare un input fisso o simulato e fare clic su **Analisi di sensibilità**. L'analisi di sensibilità è limitata a un solo input fisso o a un solo parametro di distribuzione per un input simulato. Per ulteriori informazioni, consultare l'argomento “Analisi di sensibilità ” a pagina 297.

Icone dello stato di adattamento

Le icone nella colonna Adatta a, indicano lo stato di adattamento per ciascun campo di input.

Tabella 8. Icone di stato

Icona	Descrizione
	Non è stata specificata alcuna distribuzione per l'input e l'input non è stato specificato come fisso. Per eseguire la simulazione, è necessario specificare una distribuzione per questo input o definirlo come fisso e specificare il valore fisso.
	L'input è stato adattato in precedenza a un campo che non esiste nel dataset attivo. Non è necessaria alcuna azione, a meno che non si desideri riadattare la distribuzione per l'input al dataset attivo.
	La distribuzione dell'adattamento più vicina è stata sostituita con una distribuzione alternativa dalla finestra di dialogo Dettagli adattamento.
	L'input viene impostato sulla distribuzione dell'adattamento più vicina.
	La distribuzione è stata specificata manualmente o le iterazioni dell'analisi di sensibilità sono state specificate per questo input.

Dettagli adattamento

Nella finestra di dialogo Dettagli adattamento vengono visualizzati i risultati dell'adattamento della distribuzione automatica per un particolare input. Le distribuzioni sono ordinate per bontà di adattamento, per cui la distribuzione dell'adattamento più vicino viene elencata per prima. È possibile sovrascrivere la distribuzione dell'adattamento più vicina selezionando il pulsante di scelta per la distribuzione desiderata nella colonna Utilizza. Selezionando un pulsante di scelta nella colonna Utilizza, viene visualizzato anche un grafico della distribuzione sovrapposto ad un istogramma (o grafico a barre) dei dati cronologici per l'input in questione.

Statistiche di adattamento. Per impostazione predefinita, per i campi continui il test di Anderson-Darling viene utilizzato per determinare la bontà dell'adattamento. In alternativa, solo per i campi continui, è possibile specificare il test di Kolmogorov-Smirnoff per la bontà dell'adattamento selezionando tale opzione nelle impostazioni di Opzioni avanzate. Per gli input continui, i risultati di entrambi i test vengono visualizzati nella colonna Statistiche di adattamento (A per Anderson-Darling e K per Kolmogorov-Smirnoff) e il test scelto viene utilizzato per ordinare le distribuzioni. Per gli input ordinali e nominali viene utilizzato il test del chi-quadrato. Inoltre, vengono visualizzati i valori p associati ai test.

Parametri. I parametri della distribuzione associati a ogni distribuzione adattata vengono visualizzati nella colonna Parametri. I parametri per le seguenti distribuzioni hanno un significato identico a quello nelle funzioni delle variabili casuali associate disponibili nella finestra di dialogo Calcola variabile: Bernoulli, beta, binomiale, esponenziale, gamma, lognormale, binomiale negativa (prove ed errori), normale, Poisson e uniforme. Per la distribuzione categoriale, i nomi dei parametri sono le categorie e i valori dei parametri sono le probabilità associate.

Riadattamento con un insieme di distribuzioni personalizzate. Per impostazione predefinita, il livello di misurazione dell'input viene utilizzato per determinare l'insieme di distribuzioni considerate per l'adattamento della distribuzione automatica. Ad esempio, le distribuzioni continue, quali quelle lognormali e gamma, vengono considerate durante l'adattamento di un input continuo, ma le distribuzioni discrete, quali quelle Poisson e binomiali, non vengono considerate. È possibile scegliere un sottoinsieme delle distribuzioni predefinite selezionando le distribuzioni nella colonna Riadatta. Inoltre, è possibile sovrascrivere l'insieme predefinito di distribuzioni, selezionando un livello di misurazione differente nell'elenco a discesa **Tratta come (misura)** e scegliendo le distribuzioni nella colonna Riadatta. Fare clic su **Esegui riadattamento** per eseguire il riadattamento con l'insieme di distribuzioni personalizzate.

Nota:

- I casi con valori mancanti per eventuale input simulato vengono esclusi dall'adattamento della distribuzione, dal calcolo delle correlazioni e dal calcolo della tabella di contingenza facoltativa (per input con una distribuzione categoriale). È possibile, facoltativamente, specificare se i valori mancanti definiti dall'utente di input con una distribuzione categoriale sono trattati come validi. Per impostazione predefinita, vengono trattati come mancanti. Per ulteriori informazioni, consultare l'argomento “Opzioni avanzate ” a pagina 298.
- Per gli input continui e ordinali, se non viene trovato un adattamento accettabile per una qualsiasi delle distribuzioni testate, la distribuzione empirica viene suggerita come adattamento più vicino. Per gli input continui, la distribuzione empirica è la funzione di distribuzione cumulativa dei dati cronologici. Per gli input ordinali, la distribuzione empirica è la distribuzione categoriale dei dati cronologici.

Analisi di sensibilità

L'analisi della sensibilità consente di esaminare l'effetto che si ottiene modificando un input fisso o un parametro di distribuzione per un input simulato in un insieme specificato di valori. Un insieme indipendente di casi simulati - in effetti, una simulazione separata - viene generato per ogni valore specificato, consentendo all'utente di esaminare l'effetto della modifica dell'input. Ogni insieme di casi simulati viene definito come un'**iterazione**.

Itera. Questa opzione consente di specificare l'insieme di valori sui quali verrà modificato l'input.

- Se si modifica il valore di un parametro di distribuzione, selezionare il parametro dall'elenco a discesa. Immettere l'insieme di valori nella griglia Valore parametro per iterazione. Facendo clic su **Continua**, i valori specificati verranno aggiunti alla griglia Parametri per l'input associato e un indice specifica il numero di iterazioni del valore.
- Per le distribuzioni Categoriale e Intervallo, è possibile modificare le probabilità rispettivamente delle categorie o degli intervalli, ma i valori delle categorie e degli endpoint degli intervalli non possono essere modificati. Selezionare una categoria o un intervallo dall'elenco a discesa e specificare l'insieme di probabilità nella griglia Valore parametro per iterazione. Le probabilità per le altre categorie o gli altri intervalli verranno adattate automaticamente.

Nessuna iterazione. Utilizzare questa opzione per annullare le iterazioni per un input. Facendo clic su **Continua**, le iterazioni verranno rimosse.

Correlazioni

I campi di input da simulare sono spesso notoriamente correlati, ad esempio altezza e peso. Le correlazioni tra gli input che verranno simulati devono essere prese in considerazione per essere certi che i valori simulati conservino tali correlazioni.

Ricalcola correlazioni durante l'adattamento. Questa opzione specifica che le correlazioni tra gli input simulati vengono calcolate automaticamente durante l'adattamento delle distribuzione al dataset attivo mediante le azioni **Adatta tutto** o **Adatta** nelle impostazioni di Campi simulati.

Non ricalcolare correlazioni durante l'adattamento. Selezionare questa opzione se si desidera specificare manualmente le correlazioni e impedire che vengano sovrascritte durante l'adattamento automatico delle distribuzioni al dataset attivo. I valori inseriti nella griglia delle correlazioni devono essere compresi tra -1 e 1. Un valore di 0 specifica che non esiste alcuna correlazione tra la coppia associata di input.

Reimposta. Reimposta tutte le correlazioni su 0.

Utilizza tabella di contingenza a più vie adattata per gli input con una distribuzione categoriale. Per gli input con una distribuzione categoriale, è possibile calcolare automaticamente una tabella di contingenza a più vie dal dataset attivo che descrive le associazioni tra questi input. La tabella di contingenza viene quindi utilizzata quando i dati vengono generati per questi input. Se si sceglie di salvare il piano di simulazione, la tabella di contingenza viene salvata nel file del piano e viene utilizzata quando si esegue il piano.

- **Calcola tabella di contingenza dal dataset attivo.** Se si sta gestendo un piano di simulazione esistente che contiene una tabella di contingenza, è possibile ricalcolare la tabella di contingenza dal dataset attivo. Questa azione sovrascrive la tabella di contingenza dal file del piano caricato.
- **Utilizza tabella di contingenza dal piano di simulazione caricato.** Per impostazione predefinita, quando si carica un piano di simulazione che contiene una tabella di contingenza, viene utilizzata la tabella dal piano. È possibile ricalcolare la tabella di contingenza dal dataset attivo selezionando **Calcola tabella di contingenza dal dataset attivo**.

Opzioni avanzate

Numero massimo di casi. Specifica il numero massimo di casi dei dati simulati e i valori obiettivo associati da generare. Quando si specifica l'analisi di sensibilità, questo è il numero massimo di casi per ogni iterazione.

Destinazione per i criteri di arresto. Se il modello predittivo contiene più di un obiettivo, è possibile selezionare l'obiettivo a cui si applicano i criteri di arresto.

Criteri di arresto. Queste opzioni specificano i criteri per arrestare la simulazione, potenzialmente prima che venga generato il numero massimo di casi consentiti.

- **Continua fino al numero massimo.** Specifica che i casi simulati verranno generati finché non viene raggiunto il numero massimo di casi.
- **Arresta quando le code sono campionate.** Utilizzare questa opzione quando si desidera verificare che il campionamento di una delle code di una distribuzione obiettivo specificata sia stato eseguito in modo adeguato. I casi simulati verranno generati finché non viene completato il campionamento delle code specificate o non viene raggiunto il numero massimo di casi. Se il modello predittivo contiene più obiettivi, selezionare l'obiettivo a cui verranno applicati questi criteri, dall'elenco a discesa **Obiettivo per i criteri di arresto**.

  Tipo. È possibile definire il limite della regione della coda specificando un valore dell'obiettivo, ad esempio 10.000.000 o un percentile, ad esempio, novantanovesimo percentile. Se si sceglie Valore nell'elenco a discesa **Tipo**, immettere il valore del limite nella casella di testo Valore e utilizzare l'elenco a discesa **Lato** per specificare se si tratta del limite della regione della coda destra o sinistra. Se si sceglie Percentile nell'elenco a discesa **Tipo**, immettere un valore nella casella di testo Percentile.

  Frequenza. Specificare il numero di valori dell'obiettivo che devono trovarsi nella regione della coda per verificare che il campionamento della coda sia stato eseguito in modo adeguato. La creazione dei casi verrà arrestata quando viene raggiunto questo numero.
- **Arresta quando l'intervallo di confidenza della media rientra nella soglia specificata.** Utilizzare questa opzione quando si desidera verificare che la media di un determinato obiettivo sia nota con un grado di precisione specificato. I casi simulati verranno generati finché non viene raggiunto il grado di precisione specificato o o non viene raggiunto il numero massimo di casi. Per utilizzare questa opzione, specificare un livello di confidenza e una soglia. I casi simulati verranno generati finché l'intervallo di confidenza associato al livello specificato è all'interno della soglia. Ad esempio, è possibile utilizzare questa opzione per specificare che i casi vengono generati finché l'intervallo di confidenza della media a livello di confidenza del 95% è entro il 5% del valore della media. Se il modello predittivo contiene più obiettivi, selezionare l'obiettivo a cui verranno applicati questi criteri, dall'elenco a discesa **Obiettivo per i criteri di arresto**.

  Tipo di soglia. È possibile specificare la soglia come un valore numerico o una percentuale della media. Se si sceglie Valore nell'elenco a discesa **Tipo di soglia**, immettere la soglia nella casella di testo Soglia

come valore. Se si sceglie Percentuale nell'elenco a discesa **Tipo di soglia**, immettere un valore nella casella di testo Soglia come percentuale.

Numero di casi da campionare. Specifica il numero di casi da utilizzare per l'adattamento automatico delle distribuzioni per gli input simulati nel dataset attivo. Se il dataset è molto grande, si consiglia di limitare il numero di casi utilizzati per l'adattamento della distribuzione. Se si seleziona **Limita a N casi**, verranno utilizzati i primi N casi.

Criteri della bontà di adattamento (input continui). Per gli input continui è possibile utilizzare il test di Anderson-Darling o il test di Kolmogorov-Smirnoff della bontà dell'adattamento per classificare le distribuzioni durante il loro adattamento per gli input simulati nel dataset attivo. Il test di Anderson-Darling è selezionato per impostazione predefinita ed è particolarmente consigliato quando si desidera ottenere il migliore adattamento possibile nelle regioni delle code.

Distribuzione empirica. Per gli input continui, la distribuzione empirica è la funzione di distribuzione cumulativa dei dati cronologici. È possibile specificare il numero di bin utilizzati per calcolare la distribuzione empirica degli input continui. Il valore predefinito è 100, il valore massimo è 1000.

Replica risultati. L'impostazione di un seme casuale consente di replicare la simulazione. Specificare un numero intero o fare clic su **Genera** per creare un numero intero pseudo-casuale compreso tra 1 e 2147483647, compresi. Il valore predefinito è 629111597.

Nota: Per un seme particolarmente casuale, i risultati vengono replicati a mano che il numero di thread non sia cambiato. Su particolari computer, il numero di thread è costante a meno che non lo si cambi eseguendo la sintassi comando SET THREADS. Il numero di thread può cambiare se si esegue la simulazione su un altro computer perché viene utilizzato un algoritmo interno per determinare il numero di thread per ogni computer.

Valori mancanti definiti dall'utente per gli input con una distribuzione categoriale. Questi controlli specificano se i valori mancanti definiti dall'utente di input con una distribuzione categoriale sono trattati come validi. I valori mancanti di sistema e i valori mancanti definiti dall'utente per tutti gli altri tipi di input sono sempre trattati come non validi. Tutti gli input devono avere dei valori validi per un caso per essere inclusi nell'adattamento della distribuzione, nel calcolo delle correlazioni e nel calcolo della tabella di contingenza facoltativa.

Funzioni di densità

Queste impostazioni consentono di personalizzare l'output per le funzioni di densità di probabilità e le funzioni di distribuzione cumulativa per gli obiettivi continui, nonché i grafici a barre dei valori previsti per gli obiettivi categoriali.

Funzione di densità di probabilità (PDF). La funzione di densità di probabilità visualizza la distribuzione dei valori obiettivo. Per gli obiettivi continui consente di determinare la probabilità che la destinazione sia all'interno di una regione data. Per gli obiettivi categoriali (obiettivi con un livello di misurazione nominale o ordinale), viene generato un grafico a barre che visualizza la percentuale di casi che rientrano in ciascuna categoria di obiettivo. Altre opzioni per gli obiettivi categoriali dei modelli PMML sono disponibili con l'impostazione Valori di categoria da segnalare descritta di seguito.

Per i modelli di cluster Two-Step e delle k-medie, viene generato un grafico a barre dell'appartenenza al cluster.

Funzione di distribuzione cumulativa (CDF). La funzione di distribuzione cumulativa visualizza la probabilità che il valore dell'obiettivo sia inferiore o uguale a un valore specificato. È disponibile solo per gli obiettivi continui.

Posizioni dispositivo di scorrimento. È possibile specificare le posizioni iniziali delle linee di riferimento mobili su grafici PDF o CDF. I valori specificati per le linee inferiore e superiore fanno riferimento a posizioni lungo l'asse orizzontale, non a percentili. È possibile rimuovere la linea inferiore selezionando **-Infinità** oppure è possibile rimuovere la linea superiore selezionando **Infinità**. Per impostazione predefinita, le linee sono posizionate al 5° e al 95° percentile. Quando vengono visualizzate più funzioni di distribuzione su un singolo grafico (a causa di più obiettivi o risultati dalle iterazioni dell'analisi di sensibilità), il valore predefinito fa riferimento alla distribuzione per la prima iterazione o il primo obiettivo.

Linee di riferimento (continue). È possibile richiedere diverse linee di riferimento verticali da aggiungere alle funzioni di densità di probabilità e alle funzioni di distribuzione cumulativa per gli obiettivi continui.

- **Sigma.** È possibile aggiungere le linee di riferimento a più e meno di un numero specificato di deviazioni standard dalla media di un obiettivo.
- **Percentili.** È possibile aggiungere delle linee di riferimento a uno o due valori di percentile della distribuzione di un obiettivo, immettendo i valori nelle caselle di testo Basso e Alto. Ad esempio, il valore 95 nella casella di testo Alto rappresenta il 95° percentile, ovvero il valore al di sotto del quale ricade il 95% delle osservazioni. Allo stesso modo, il valore 5 nella casella di testo Basso rappresenta il 5° percentile, ovvero il valore al di sotto del quale ricade il 5% delle osservazioni.
- **Linee di riferimento personalizzate.** È possibile aggiungere le linee di riferimento nei valori specificati dell'obiettivo.

Nota: Quando vengono visualizzate più funzioni di distribuzione su un singolo grafico (a causa di più obiettivi o risultati dalle iterazioni dell'analisi di sensibilità), le linee di riferimento vengono applicate solo alla distribuzione per la prima iterazione o per il primo obiettivo. È possibile aggiungere delle linee di riferimento alle altre distribuzioni dalla finestra di dialogo Grafico: opzioni, a cui si accede dal grafico PDF o CDF.

Sovrapponi risultati dagli obiettivi continui separati. Nel caso di più obiettivi continui, specifica se le funzioni di distribuzione per tutti gli obiettivi vengono visualizzate in un solo grafico, con un grafico per le funzioni di densità di probabilità e un altro per le funzioni di distribuzione cumulativa. Quando questa opzione non è selezionata, i risultati di ciascun obiettivo verranno visualizzati in un grafico separato.

Valori di categoria da segnalare. Per i modelli PMML con destinazioni categoriali, il risultato del modello è un insieme di probabilità previste, una per ogni categoria, che il valore obiettivo rientri in ogni categoria. La categoria con la probabilità più alta viene considerata come la categoria prevista e viene utilizzata per la creazione del grafico a barre descritto per l'impostazione **Funzione di densità di probabilità** descritta in precedenza. Selezionando **Categoria prevista**, verrà generato il grafico a barre. Selezionando **Probabilità previste**, verranno generati gli istogrammi della distribuzione delle probabilità previste per ciascuna delle categorie di obiettivo.

Raggruppamento per analisi di sensibilità. Le simulazioni che includono l'analisi di sensibilità generano un insieme indipendente di valori obiettivo previsti per ogni iterazione definita dall'analisi (un'iterazione per ogni valore dell'input che viene modificato). Quando sono presenti le iterazioni, il grafico a barre della categoria prevista per un obiettivo categoriale viene visualizzato come un grafico a barre raggruppate che include i risultati di tutte le iterazioni. È possibile scegliere di raggruppare le categorie o le iterazioni.

Output

Grafici tornado. I grafici tornado sono dei grafici a barre che visualizzano le relazioni tra gli obiettivi e gli input simulati utilizzando una varietà di metriche.

- **Correlazione dell'obiettivo con l'input.** Questa opzione crea un grafico tornado dei coefficienti di correlazioni tra un obiettivo dato e ognuno dei suoi input simulati. Questo tipo di grafico tornado non supporta gli obiettivi con un livello di misurazione nominale od ordinale oppure input simulati con una distribuzione categoriale.
- **Contributo alla varianza.** Questa opzione crea un grafico tornado che visualizza il contributo della varianza di un obiettivo da ognuno dei suoi input simulati, consentendo all'utente di valutare in che misura ciascun input contribuisce all'incertezza globale nell'obiettivo. Questo tipo di grafico tornado non supporta gli obiettivi con livelli di misurazione ordinali o nominali o input simulati con una qualsiasi delle seguenti distribuzioni: categoriale, Bernoulli, binomiale, Poisson o binomiale negativa.
- **Sensibilità dell'obiettivo al cambiamento.** Questa opzione crea un grafico tornado che visualizza l'effetto sull'obiettivo della modulazione di ogni input simulato per più o meno un numero specificato di deviazioni standard della distribuzione associata all'input. Questo tipo di grafico tornado non supporta gli obiettivi con livelli di misurazione ordinali o nominali o input simulati con una qualsiasi delle seguenti distribuzioni: categoriale, Bernoulli, binomiale, Poisson o binomiale negativa.

Grafici a scatole delle distribuzioni obiettivo. I grafici a scatole sono disponibili per gli obiettivi continui. Selezionare **Sovrapponi risultati da obiettivi separati** se i modelli predittivi hanno più obiettivi continui e si desidera visualizzare i grafici a scatole per tutti gli obiettivi in un solo grafico.

Grafici a dispersione degli obiettivi rispetto agli input. I grafici a dispersione degli obiettivi rispetto agli input simulati sono disponibili sia per gli obiettivi continui che per quelli categoriali e includono i grafici a dispersione dell'obiettivo con input continui e categoriali. Le dispersioni che interessano un obiettivo categoriale o un input categoriale sono visualizzate come una mappa termica.

Crea una tabella dei valori percentili. Per gli obiettivi continui è possibile ottenere una tabella dei percentili specificati delle distribuzioni di destinazione. I quartili (il 25°, 50° e 75° percentile) suddividono le osservazioni in quattro gruppi di dimensioni uguali. Se si desidera avere un numero uguale di gruppi diverso da quattro, selezionare **Intervalli** e specificare il numero. Selezionare **Percentili personalizzati** per specificare i singoli percentili, ad esempio il 99esimo percentile.

Statistiche descrittive delle distribuzioni obiettivo. Questa opzione crea le tabelle di statistiche descrittive per gli obiettivi continui e categoriali, nonché per gli input continui. Per gli obiettivi continui la tabella include la media, la deviazione standard, la mediana, i valori minimo e massimo, l'intervallo di confidenza della media nel livello specificato e il quinto e 95esimo percentile della distribuzione di destinazione. Per gli obiettivi categoriali la tabella include la percentuale di casi che rientrano in ciascuna categoria dell'obiettivo. Per gli obiettivi categoriali dei modelli PMML la tabella include anche la probabilità della media di ciascuna categoria obiettivo. Per gli input continui la tabella include la media, la deviazione standard e i valori minimo e massimo.

Correlazioni e tabella di contingenza per gli input. Questa opzione visualizza una tabella di coefficienti di correlazione tra gli input simulati. Quando gli input con le distribuzioni categoriali sono generati da una tabella di contingenza, viene visualizzata anche la tabella di contingenza dei dati generati per questi input.

Input simulati da includere nell'output. Per impostazione predefinita, tutti gli input simulati sono inclusi nell'output. È possibile escludere gli input simulati selezionati dall'output. In tal modo, verranno esclusi dai grafici tornado, dai grafici a dispersione e dall'output tabulare.

Intervalli di limite per gli obiettivi continui. È possibile specificare l'intervallo di valori validi per uno o più obiettivi continui. I valori fuori dall'intervallo specificato sono esclusi da tutto l'output e da tutte le analisi associate agli obiettivi. Per impostare un limite inferiore, selezionare **Inferiore** nella colonna Limite e immettere un valore nella colonna Minimo. Per impostare un limite superiore, selezionare **Superiore** nella colonna Limite e immettere un valore nella colonna Massimo. Per impostare sia un limite inferiore che un limite superiore, selezionare **Entrambi** nella colonna Limite e immettere i valori nelle colonne Minimo e Massimo.

Formati di visualizzazione. È possibile impostare il formato utilizzato durante la visualizzazione dei valori degli obiettivi e degli input (input fissi e input simulati).

Salvataggio di

Salva il file del piano per questa simulazione. È possibile salvare le specifiche correnti per la simulazione in un file del piano di simulazione. I file del piano di simulazione hanno estensione *.splan*. È possibile riaprire il piano nel Builder di simulazioni, apportare le modifiche, se lo si desidera, ed eseguire la simulazione. È possibile condividere il piano di simulazione con altri utenti, che possono quindi eseguirla nella finestra di dialogo Esegui simulazione. I piani di simulazione includono tutte le specifiche tranne le seguenti: impostazioni per Funzioni di densità; impostazioni di Output per i grafici e le tabelle; impostazioni di Opzioni avanzate per Adattamento, Distribuzione empirica e Seme casuale.

Salva i dati simulati come nuovo file di dati. È possibile salvare gli input simulati, gli input fissi e i valori obiettivo previsti in un file di dati SPSS Statistiche, in un nuovo dataset nella sessione corrente o in un file Excel. Ogni caso (o riga) del file di dati è composto dai valori previsti delle destinazioni insieme agli input simulati e fissi che generano i valori obiettivo. Quando si specifica l'analisi di sensibilità, ogni iterazione dà luogo a un insieme continuo di casi che sono etichettati con il numero dell'iterazione.

Finestra di dialogo Esegui simulazione

La finestra di dialogo Esegui simulazione è progettata per gli utenti che hanno un piano di simulazione e desiderano principalmente eseguire una simulazione. Essa offre anche le funzioni necessarie per eseguire la simulazione in condizioni differenti. Consente di eseguire le attività generali riportate di seguito.

- Impostare o modificare l'analisi di sensibilità per esaminare l'effetto che si ottiene modificando il valore di un input fisso o modificando un parametro di distribuzione di un input simulato.
- Riadattare le distribuzioni della probabilità per gli input incerti (e le correlazioni tra tali input) ai nuovi dati.
- Modificare la distribuzione di un input simulato.
- Personalizzare l'output.
- Eseguire la simulazione.

Scheda Simulazione

La scheda Simulazione consente di specificare l'analisi di sensibilità, riadattare le distribuzioni di probabilità per gli input simulati e le correlazioni tra gli input simulati e i nuovi dati, nonché modificare la distribuzione della probabilità associata a un input simulato.

La griglia Input simulati contiene una voce per ogni campo di input definito nel piano di simulazione. Ogni voce visualizza il nome dell'input e il tipo di distribuzione della probabilità associato all'input, insieme a un grafico di esempio della curva di distribuzione associata. Inoltre, a ogni input è associata un'icona di stato (un cerchio colorato con un segno di spunta), che è utile quando si esegue il riadattamento delle distribuzioni ai nuovi dati. Inoltre, gli input possono includere un'icona di blocco, che indica che l'input è bloccato e non può essere modificato o riadattato ai nuovi dati nella finestra di dialogo Esegui simulazione. Per modificare un input bloccato, è necessario aprire il piano di simulazione nel Builder di simulazioni.

Ogni input è simulato o fisso. Gli input simulati sono quelli i cui valori sono incerti e verranno generati a partire da una distribuzione della probabilità specificata. Gli input fissi sono quelli i cui valori sono noti e rimangono costanti per ogni caso generato nella simulazione. Per utilizzare un determinato input, selezionare la voce relativa all'input nella griglia Input simulati.

Specifica dell'analisi di sensibilità

L'analisi di sensibilità consente di investigare l'effetto delle modifiche sistematiche in un input fisso o in un parametro di distribuzione per un input simulato generando un insieme indipendente di casi simulati - in modo efficace, una simulazione separata - per ogni valore specificato. Per specificare l'analisi di sensibilità, selezionare un input fisso o simulato e fare clic su **Analisi di sensibilità**. L'analisi di sensibilità è limitata a un solo input fisso o a un solo parametro di distribuzione per un input simulato. Per ulteriori informazioni, consultare l'argomento “Analisi di sensibilità ” a pagina 297.

Riadattamento delle distribuzioni ai nuovi dati

Per riadattare automaticamente le distribuzioni della probabilità per gli input simulati (e le correlazioni tra gli input simulati) ai dati nel dataset attivo, effettuare quanto segue.

1. Verificare che ciascuno degli input del modello corrisponda al campo corretto nel dataset attivo. Ogni input simulato è adattato al campo nel dataset attivo specificato nell'elenco a discesa **Campo** associato all'input in questione. È possibile identificare facilmente gli input che non hanno corrispondenza, cercando gli input con un'icona di stato che include un segno di spunta con un punto interrogativo, come illustrato di seguito.

2. Modificare le corrispondenze dei campi necessarie selezionando **Adatta a un campo nel dataset**, quindi selezionando il campo nell'elenco.
3. Fare clic su **Adatta tutto**.

Per ogni input che è stato adattato, viene visualizzata la distribuzione che si adatta meglio ai dati, insieme a un grafico della distribuzione sovrapposto a un istogramma (o grafico a barre) dei dati cronologici. Se non viene trovato un adattamento accettabile, viene utilizzata la distribuzione empirica. Per gli input che sono adattati alla distribuzione empirica, viene visualizzato solo un istogramma dei dati cronologici perché la distribuzione empirica è infatti rappresentata da tale istogramma.

Nota: per un elenco completo di icone di stato, consultare l'argomento "Campi simulati" a pagina 294.

Modifica delle distribuzioni della probabilità

È possibile modificare la distribuzione della probabilità per un input simulato e, facoltativamente, modificare un input simulato in un input fisso o viceversa.

1. Selezionare l'input, quindi selezionare **Imposta manualmente la distribuzione**.
2. Selezionare il tipo di distribuzione e specificare i parametri della distribuzione. Per modificare un input simulato in un input fisso, selezionare Fisso nell'elenco a discesa **Tipo**.

Dopo avere immesso i parametri di una distribuzione, il grafico di esempio della distribuzione (visualizzato nell'immissione dell'input) viene aggiornato per riflettere le modifiche. Per ulteriori informazioni su come specificare manualmente le distribuzioni della probabilità, consultare l'argomento "Campi simulati" a pagina 294.

Includi valori mancanti definiti dall'utente degli input categoriali in fase di adattamento. Specifica se i valori mancanti definiti dall'utente di input con una distribuzione categoriale sono trattati come validi quando si sta eseguendo il riadattamento ai nel dataset attivo. I valori mancanti di sistema e i valori mancanti definiti dall'utente per tutti gli altri tipi di input sono sempre trattati come non validi. Tutti gli input devono avere dei valori validi per un caso per essere inclusi nell'adattamento della distribuzione e nel calcolo delle correlazioni.

Scheda Output

La scheda Output consente di personalizzare l'output generato dalla simulazione.

Funzioni di densità. Le funzioni di densità sono lo strumento principale per sondare l'insieme di risultati ottenuti dalla simulazione.

- **Funzione di densità di probabilità.** La funzione di densità di probabilità visualizza la distribuzione dei valori obiettivo, consentendo di determinare la probabilità che la destinazione sia all'interno di una regione data. Per obiettivi con un insieme fisso di risultati, ad esempio "servizio insufficiente", "servizio sufficiente", "servizio buono" e "servizio eccellente", viene generato un grafico a barre che visualizza la percentuale di casi che rientrano in ciascuna categoria di obiettivo.
- **Funzione di distribuzione cumulativa.** La funzione di distribuzione cumulativa visualizza la probabilità che il valore dell'obiettivo sia inferiore o uguale a un valore specificato.

Grafici tornado. I grafici tornado sono dei grafici a barre che visualizzano le relazioni tra gli obiettivi e gli input simulati utilizzando una varietà di metriche.

- **Correlazione dell'obiettivo con l'input.** Questa opzione crea un grafico tornado dei coefficienti di correlazioni tra un obiettivo dato e ognuno dei suoi input simulati.
- **Contributo alla varianza.** Questa opzione crea un grafico tornado che visualizza il contributo della varianza di un obiettivo da ognuno dei suoi input simulati, consentendo all'utente di valutare in che misura ciascun input contribuisce all'incertezza globale nell'obiettivo.
- **Sensibilità dell'obiettivo al cambiamento.** Questa opzione crea un grafico tornado che visualizza l'effetto sull'obiettivo della modulazione di ogni input simulato per più o meno una deviazione standard della distribuzione associata all'input.

Grafici a dispersione degli obiettivi rispetto agli input. Questa opzione genera dei grafici a dispersione degli obiettivi rispetto agli input simulati.

Grafici a scatole delle distribuzioni obiettivo. Questa opzione genera dei grafici a scatole delle distribuzioni obiettivo.

Tabella quartili. Questa opzione genera una tabella dei quartili delle distribuzioni obiettivo. I quartili di una distribuzione sono il 25esimo, il 50esimo e il 75esimo percentile della distribuzione e dividono le osservazioni in quattro gruppi della stessa dimensione.

Correlazioni e tabella di contingenza per gli input. Questa opzione visualizza una tabella di coefficienti di correlazione tra gli input simulati. Una tabella di contingenza di associazioni tra gli input con una distribuzione categoriale viene visualizzata quando il piano di simulazione specifica la generazione di dati categoriali da una tabella di contingenza.

Sovrapponi risultati da obiettivi separati. Se il modello predittivo che si sta simulando contiene più obiettivi, è possibile specificare se i risultati ottenuti da obiettivoi separati vengono visualizzati in un solo grafico. Questa impostazione si applica alle funzioni di densità di probabilità, alle funzioni di distribuzione cumulativa e ai grafici a scatole. Ad esempio, se si seleziona questa opzione, le funzioni di densità di probabilità per tutti gli obiettivi verranno visualizzate in un solo grafico.

Salva il file del piano per questa simulazione. È possibile salvare le modifiche apportate alla simulazione in un file del piano di simulazione. I file del piano di simulazione hanno estensione *.splan*. È possibile riaprire il piano nella finestra di dialogo Esegui simulazione o nel Builder di simulazioni. I piani di simulazione includono tutte le specifiche, tranne le impostazioni di output.

Salva i dati simulati come nuovo file di dati. È possibile salvare gli input simulati, gli input fissi e i valori obiettivo previsti in un file di dati SPSS Statistiche, in un nuovo dataset nella sessione corrente o in un file Excel. Ogni caso (o riga) del file di dati è composto dai valori previsti delle destinazioni insieme agli input simulati e fissi che generano i valori obiettivo. Quando si specifica l'analisi di sensibilità, ogni iterazione dà luogo a un insieme continuo di casi che sono etichettati con il numero dell'iterazione.

Se è necessaria una maggiore personalizzazione dell'output di quella disponibile qui, si consiglia di eseguire la simulazione dal Builder di simulazioni. Per ulteriori informazioni, consultare l'argomento “Eseguire una simulazione da un piano di simulazione” a pagina 290.

Utilizzo dell'output del grafico dalla simulazione

Numerosi grafici generati da una simulazione hanno funzioni interattive che consentono di personalizzare la visualizzazione. Le funzioni interattive sono disponibili attivando (facendo doppio clic) l'oggetto grafico nel Visualizzatore di output. Tutti i grafici di simulazione sono visualizzazioni di tipo lavagna grafica.

Grafici Funzione di densità di probabilità per gli obiettivi continui. Questo grafico ha due righe di riferimento verticali scorrevoli che dividono il grafico in regioni separate. La tabella sotto il grafico mostra la probabilità che l'obiettivo sia in ognuna delle regioni. Se più funzioni di densità vengono visualizzate nello stesso grafico, la tabella ha una riga separata per le probabilità associate a ciascuna funzione di densità. Ognuna delle righe di riferimento ha un dispositivo di scorrimento (triangolo invertito) che consente di spostare facilmente la linea. Sono disponibili numerose altre funzioni facendo clic sul pulsante **Grafici: opzioni** nel grafico. In particolare, è possibile impostare esplicitamente le posizioni dei dispositivi di scorrimento, aggiungere righe di riferimento fisse e modificare la vista del grafico da una curva continua in un istogramma o viceversa. Per ulteriori informazioni, consultare l'argomento “Grafico: opzioni ” a pagina 305.

Grafici Funzione di distribuzione cumulativa per gli obiettivi continui. Questo grafico ha le stesse due righe di riferimento verticali scorrevoli e la tabella associata descritte per il grafico Funzione di densità di probabilità illustrato in precedenza. Inoltre, consente di accedere alla finestra di dialogo Grafici: opzioni, in cui è possibile impostare esplicitamente le posizioni dei dispositivi di scorrimento, aggiungere righe di riferimento fisse e specificare se la funzione di distribuzione cumulativa viene visualizzata come una funzione crescente (impostazione predefinita) o come una funzione decrescente. Per ulteriori informazioni, consultare l'argomento “Grafico: opzioni ” a pagina 305.

Grafici a barre per gli obiettivi categoriali con iterazioni dell'analisi di sensibilità. Per le destinazioni categoriali con iterazioni dell'analisi di sensibilità i risultati per la categoria obiettivo prevista vengono visualizzati come un grafico a barre raggruppate, che include i risultati di tutte le iterazioni. Il grafico include un elenco a discesa che consente di eseguire il raggruppamento sulla categoria o sull'iterazione. Per i modelli di cluster Two-Step e delle k-medie, è possibile scegliere di eseguire il raggruppamento sul numero di cluster o sull'iterazione.

Grafici a scatole per più obiettivi con iterazioni dell'analisi di sensibilità. Per i modelli predittivi con più obiettivi continui e iterazioni dell'analisi di sensibilità, se si sceglie di visualizzare i grafici a scatole per tutti gli obiettivi in un solo grafico, viene generato un grafico a scatole raggruppate. Il grafico include un elenco a discesa che consente di eseguire il raggruppamento sull'obiettivo o sull'iterazione.

Grafico: opzioni

La finestra di dialogo Grafici: opzioni consente di personalizzare la visualizzazione dei grafici attivati delle funzioni di densità di probabilità e delle funzioni di distribuzione cumulativa generate da una simulazione.

Visualizza. L'elenco a discesa **Visualizza** si applica solo al grafico Funzione di densità di probabilità. Consente di alternare la visualizzazione del grafico da una curva continua in un istogramma. Questa funzione non è disponibile quando vengono visualizzate più funzioni di densità nello stesso grafico. In tal caso, le funzioni di densità possono solo essere visualizzate come curve continue.

Ordina. L'elenco a discesa **Ordina** si applica solo al grafico Funzione di distribuzione cumulativa. Specifica se la funzione di distribuzione cumulativa viene visualizzata come una funzione crescente (impostazione predefinita) o come una funzione decrescente. Quando viene visualizzata come una funzione decrescente, il valore della funzione in un determinato punto sull'asse orizzontale è la probabilità che l'obiettivo si trovi a destra di tale punto.

Posizioni dispositivo di scorrimento. È possibile impostare esplicitamente le posizioni delle righe di riferimento scorrevoli immettendo i valori nelle caselle di testo Superiore e Inferiore. È possibile rimuovere la linea a sinistra selezionando **-Infinito**, impostando in modo efficace la posizione su infinito negativo, ed è possibile rimuovere la linea a destra selezionando **Infinito**, impostando in modo efficace la sua posizione su infinito.

Righe di riferimento. È possibile aggiungere diverse linee di riferimento verticali fisse alle funzioni di densità di probabilità e alle funzioni di distribuzione cumulativa. Quando su un singolo grafico vengono visualizzate più funzioni (a causa di più obiettivi o risultati da iterazioni dell'analisi di sensibilità), è possibile specificare le specifiche funzioni a cui vengono applicate le linee.

- **Sigma.** È possibile aggiungere le linee di riferimento a più e meno di un numero specificato di deviazioni standard dalla media di un obiettivo.
- **Percentili.** È possibile aggiungere delle linee di riferimento a uno o due valori di percentile della distribuzione di un obiettivo, immettendo i valori nelle caselle di testo Basso e Alto. Ad esempio, il valore 95 nella casella di testo Alto rappresenta il 95° percentile, ovvero il valore al di sotto del quale ricade il 95% delle osservazioni. Allo stesso modo, il valore 5 nella casella di testo Basso rappresenta il 5° percentile, ovvero il valore al di sotto del quale ricade il 5% delle osservazioni.
- **Posizioni personalizzate.** È possibile aggiungere le righe di riferimento nei valori specificati lungo l'asse orizzontale.

Linee di riferimento etichetta. Questa opzione controlla se le etichette sono applicate alle linee di riferimento selezionate.

Le linee di riferimento vengono rimosse deselezionando l'opzione associata nella finestra di dialogo Grafici: opzioni e facendo clic su **Continua**.

Modellazione geospaziale

Le tecniche di modellazione geospaziale sono progettate per rilevare i modelli nei dati che includono una componente geospaziale (mappa). La procedura guidata Modellazione geospaziale fornisce i metodi per l'analisi dei dati geospaziali con e senza un componente temporale.

Trovare associazioni in base a dati evento e geospaziali (regole di associazione geospaziale)
Utilizzando le regole di associazione geospaziale, è possibile trovare i modelli nei dati in base a proprietà sia spaziali che non spaziali. Ad esempio, è possibile identificare i modelli nei dati di un crimine in base agli attributi demografici e di ubicazione. Da tali modelli è possibile creare regole che prevedano dove è probabile che si verifichino alcuni tipi di crimine.

Effettuare previsioni utilizzando serie storiche e dati geospaziali (previsione spazio-temporale)
La previsione temporale spaziale utilizza i dati che contengono dati di ubicazione, campi di input per la previsione (predittori), uno o più campi temporali e un campo obiettivo. Ogni ubicazione ha numerose righe nei dati che rappresentano i valori di ciascun predittore e l'obiettivo ad ogni intervallo di tempo.

Utilizzo della procedura guidata Modellazione geospaziale

1. Dai menu, scegliere:

   Analizzare > **Modellazione Spaziale e temporale** > **Modelli spaziali**
2. Seguire le operazioni nella procedura guidata.

Esempi

Gli esempi dettagliati sono disponibili nel sistema della Guida.

- Regole di associazione Geospaziali: **Guida** > **Argomenti** > **Case Studies** > **Base di statistiche** > **Associazione spaziale regole**
- Previsione temporale spaziale: **Guida** > **Argomenti** > **Case Studies** > **Base di statistiche** > **Spatial temporal previsione**

Selezione di mappe

La modellazione geospaziale può utilizzare una o più origini dati di mappe. Le origini dati di mappe contengono informazioni che definiscono aree geografiche e altre caratteristiche geografiche, ad esempio strade o fiumi. Molte origini mappa contengono anche dati demografici o descrittivi di altro tipo e dati evento come report di crimini o livelli di disoccupazione. È possibile utilizzare un file di specifica della mappa precedentemente definito o definire qui le specifiche di mappa e salvarle per utilizzi futuri.

Carica una specifica di mappa
Carica un file di specifica della mappa (.mplan) precedentemente definito. Le origini dati di mappa definite qui possono essere salvate in un file di specifiche. Per la previsione temporale spaziale, se si seleziona un file di specifica della mappa che identifica più di una mappa, viene richiesto di selezionare una sola mappa dal file.

Aggiungi file mappa
Aggiungere un file forma ESRI (.shp) o archivio .zip che contiene un file forma ESRI.

- Deve essere presente un file .dbf corrispondente nella stessa ubicazione del file .shp e tale file deve avere lo stesso nome radice del file .shp.
- Se il file è un archivio .zip, i file .shp e .dbf devono avere lo stesso nome radice dell'archivio .zip.
- Se non è presente un file di proiezione (.prj) corrispondente, viene richiesto di selezionare un sistema di proiezione.

Relazione
Per le regole di associazione geospaziale, questa colonna definisce la modalità di correlazione con le funzioni nella mappa. Questa impostazione non è disponibile per la previsione temporale spaziale.

Sposta su, Sposta giù
L'ordine di strato degli elementi della mappa è determinato dall'ordine in cui appaiono nell'elenco. La prima mappa nell'elenco è l'ultimo strato.

Selezione di una mappa

Per la previsione temporale spaziale, se si seleziona un file di specifica della mappa che identifica più di una mappa, viene richiesto di selezionare una sola mappa dal file. La previsione temporale spaziale non supporta mappe multiple.

Relazione geospaziale

Per le regole di associazione geospaziale, la finestra di dialogo Relazione geospaziale definisce la modalità di correlazione con le funzioni nella mappa.

- Questa impostazione si applica solo alle regole di associazione geospaziale.
- Questa impostazione influisce solo sulle origini dati associate alle mappe specificate come dati di contesto nel passo per la selezione delle origini dati.

Relazione

Chiudi
L'evento si verifica vicino a un punto o area specificati sulla mappa.

Entro
L'evento si verifica all'interno di un'area specificata sulla mappa.

Contiene
L'area evento contiene un oggetto di contesto mappa.

Intersezioni
Le posizioni in cui le linee o le regioni da mappe differenti si intersecano tra loro.

Croce
Per mappe multiple, le posizioni in cui le linee (per strade, fiumi, linee ferroviarie) da mappe differenti si incrociano tra loro.

A nord di, A sud di, A est di, A ovest di
L'evento si verifica in un'area a nord, sud, est o ovest di un punto specificato sulla mappa.

Impostazione sistema di coordinate

Se non ci sono file di proiezione (.prj) con la mappa o se si definiscono due campi da un'origine dati come un insieme di coordinate, è necessario impostare il sistema di coordinate.

Geografico predefinito (longitudine e latitudine)
Il sistema di coordinate è costituito da longitudine e latitudine.

Cartesiano semplice (X e Y)
Il sistema di coordinate è costituito da coordinate X e Y semplici.

Utilizza un ID noto (WKID)
L'ID noto (o WKID-Well known ID) per le proiezioni comuni.

Utilizza un nome di Sistema di Coordinate
Il sistema di coordinate si basa sulla proiezione denominata. Il nome è racchiuso tra parentesi.

Impostazione della proiezione

Se il sistema di proiezione non può essere determinato dalle informazioni fornite con la mappa, è necessario specificare il sistema di proiezione. La causa più comune di questa condizione è l'assenza di un file di proiezione (.prj) associato alla mappa o il fatto che il file di proiezione non può essere utilizzato.

- **Una città, regione o paese (Mercator)**
- **Un paese vasto, più paesi o continenti (Winkel Tripel)**
- **Un'area molto vicina all'equatore (Mercator)**
- **Un'area vicina a uno dei poli (Stereografico)**

La proiezione Mercator è una proiezione comune utilizzata in molte mappe. Questa proiezione tratta il globo come un cilindro svolto su una superficie piatta. La proiezione Mercator distorce la dimensione e la forma degli oggetti grandi. Questa distorsione aumenta mano a mano che ci si allontana dall'equatore e ci si avvicina ai poli. Le proiezioni Winkel Tripel e Stereografica effettuano adattamenti per il fatto che una mappa rappresenta una porzione di una sfera tridimensionale visualizzata in due dimensioni.

Sistema di coordinate e proiezione

Se si seleziona più di una mappa e le mappe hanno sistemi di coordinate e proiezione differenti, è necessario selezionare la mappa con il sistema di proiezione che si desidera utilizzare. Tale sistema di proiezione sarà utilizzato per tutte le mappe quando combinate insieme nell'output.

Origini dati

Una origine dati può essere un file dBase fornito con il file forma, un file di dati IBM SPSS Statistica o un dataset aperto nella sessione corrente.

Dati di contesto. I dati di contesto identificano le funzioni sulla mappa. I dati di contesto possono anche contenere campi da utilizzare come input per il modello. Per utilizzare un file dBase di contesto (.dbf) associato a un file di mappa (.shp), il file dBase di contesto deve essere nella stessa ubicazione del file forma e deve avere lo stesso nome radice. Ad esempio, se il file forma è `geodata.shp`, il file dBase deve essere denominato `geodata.dbf`

Dati evento. I dati evento contengono informazioni sugli eventi che si verificano, ad esempio crimini o incidenti. Questa opzione è disponibile solo per le regole di associazione geospaziale.

Densità dei punti. L'intervallo di tempo e i dati di coordinate per le stime della densità kernel. Questa opzione è disponibile solo per la previsione temporale spaziale.

Aggiungi. Apre una finestra di dialogo per l'aggiunta delle origini dati. Una origine dati può essere un file dBase fornito con il file forma, un file di dati IBM SPSS Statistica o un dataset aperto nella sessione corrente.

Associa. Apre una finestra di dialogo per specificare gli identificativi (coordinate o chiavi) utilizzati per associare i dati alle mappe. Ogni origine dati deve contenere uno o più identificativi che associano i dati alla mappa. I file dBase contenuti in un file forma generalmente includono un campo automaticamente utilizzato come identificativo predefinito. Per altre origini dati è necessario specificare i campi utilizzati come identificativi.

Convalida chiave. Apre una finestra di dialogo per convalidare la corrispondenza della chiave tra la mappa e l'origine dati.

Regole di associazione geospaziale

- Almeno una origine dati deve essere una origine dati evento.
- Tutte le origini dati evento devono utilizzare lo stesso formato di identificativo associazione mappa: valori chiave o coordinate.
- Se le origini dati evento sono associate alle mappe con valori chiave, tutte le origini evento devono utilizzare lo stesso tipo di funzione mappa (ad es. poligoni, punti, linee).

Previsione temporale spaziale

- Deve essere presente una origine dati di contesto.
- Se esiste una sola origine dati (un file di dati senza mappe associate) deve includere valori di coordinate.
- Se sono presenti due origini dati, una deve contenere i dati di contesto e l'altra i dati di densità dei punti.
- Non è possibile includere più di due origini dati.

Aggiunta di una origine dati

Una origine dati può essere un file dBase fornito con il file forma e il file contesto, un file di dati IBM SPSS Statistica o un dataset aperto nella sessione corrente.

È possibile aggiungere la stessa origine dati più volte se si desidera utilizzare un'associazione spaziale differente con ognuna.

Associazione mappa e dati

Ogni origine dati deve contenere uno o più identificativi che associano i dati alla mappa.

Coordinate
L'origine dati contiene campi che rappresentano coordinate Cartesiane, selezionare i campi che rappresentano tali coordinate X e Y. Per le regole di associazione geospaziale, può essere presente anche una coordinata Z.

Valori chiave
I valori chiave nei campi nell'origine dati corrispondono alle chiavi mappa selezionate. Ad esempio, una mappa di regioni può avere un identificativo nome (chiave mappa) che etichetta ogni regione. Tale identificativo corrisponde a un campo nei dati che contiene anch'esso i nomi delle regioni (chiave dati). I campi sono messi in corrispondenza con le chiavi mappa in base all'ordine di visualizzazione nei due elenchi.

Convalida chiavi

La finestra di dialogo Convalida chiavi fornisce un riepilogo di corrispondenze di record tra la mappa e l'origine dati, in base alle chiavi identificativo selezionate. Se sono presenti valori chiave dei dati non corrispondenti è possibile metterli manualmente in corrispondenza con i valori chiave della mappa.

Regole di associazione geospaziale

Per le regole di associazione geospaziale, dopo la definizione delle mappe e delle origini dati occorre effettuare le seguenti operazioni nella procedura guidata:

- Se non presenti più origini dati evento, definire la modalità di unione di tali origini.
- Selezionare i campi da utilizzare come condizioni e previsioni nell'analisi.

In alternativa, è possibile anche:

- Selezionare opzioni di output differenti.
- Salvare un file modello di calcolo del punteggio.
- Creare nuovi campi per i valori previsti e le regole nelle origini dati utilizzate nel modello.
- Personalizzare le impostazioni per la creazione di regole di associazione.
- Personalizzare le impostazioni di raccolta e aggregazione.

Definizione campi dati evento

Per le regole di associazione geospaziale, se esiste più di una origine dati evento, le origini vengono unite.

- Per impostazione predefinita, sono inclusi solo i campi che sono in comune tra tutte le origini dati evento.
- È possibile visualizzare un elenco di campi comuni, campi per una specifica origine dati o campi da tutte le origini dati e selezionare i campi che si desidera includere.
- Per i campi comuni, i valori **Tipo** e **Misurazione** devono essere uguali per tutte le origini dati. Se sono presenti conflitti, è possibile specificare il tipo e il livello di misurazione da utilizzare per ciascun campo comune.

Selezione campi

L'elenco di campi disponibili include i campi dalle origini dati evento e i campi dalle origini dati contesto.

- È possibile controllare l'elenco dei campi visualizzati selezionando un'origine dati dall'elenco **Origini dati**.
- È necessario selezionare almeno due campi. Almeno uno deve essere una condizione e almeno uno deve essere una previsione. Esistono diversi modi per rispettare questi requisiti, incluso la selezione di due campi per l'elenco **Entrambe (Condizione e Previsione)**.

- Le regole di associazione prevedono i valori dei campi di previsione basati sui valori dei campi di condizione. Ad esempio, nella regola "If x=1 and y=2, then z=3", i valori di x e y sono condizioni e il valore di z è la previsione.

Output

Tabelle di regole
Ogni tabella di regole visualizza i valori e le regole alle prime posizioni per confidenza, supporto regola, guadagno cumulativo, supporto condizione e distribuibilità. Ogni tabella è ordinata in base ai valori del criterio selezionato. È possibile visualizzare tutte le regole o il **Numero** di regole migliori, in base al criterio selezionato.

Cloud di termini ordinabili
Un elenco di regole migliori, in base ai valori del criterio selezionato. La dimensione del testo indica l'importanza relativa della regola. L'oggetto output interattivo contiene le regole alle prime posizioni per confidenza, supporto regola, guadagno cumulativo, supporto condizione e distribuibilità. Il criterio selezionato determina quale elenco di regole viene visualizzato per impostazione predefinita. È possibile selezionare un criterio differente interattivamente nell'output. Il valore **Numero massimo di regole da visualizzare** determina il numero di regole visualizzate nell'output.

Mappa
Grafico a barre interattivo e mappa delle regole migliori, in base al criterio selezionato. Ogni oggetto output interattivo contiene le regole alle prime posizioni per confidenza, supporto regola, guadagno cumulativo, supporto condizione e distribuibilità. Il criterio selezionato determina quale elenco di regole viene visualizzato per impostazione predefinita. È possibile selezionare un criterio differente interattivamente nell'output. Il valore **Numero massimo di regole da visualizzare** determina il numero di regole visualizzate nell'output.

Tabelle di informazioni sul modello

Trasformazioni campo.
Descrive le trasformazioni applicate ai campi utilizzati nell'analisi.

Riepilogo record.
Numero e percentuale di record inclusi ed esclusi.

Statistiche regola.
Statistiche di riepilogo per supporto condizione, confidenza, supporto regola, guadagno cumulativo e distribuibilità. Le statistiche includono media, minimo, massimo e deviazione standard.

Elementi più frequenti.
Elementi che ricorrono più frequentemente. Un elemento è incluso in una condizione o una previsione in una regola. Ad esempio, età < 18 o genere=femmina.

Campi più frequenti.
I campi che ricorrono più frequentemente nelle regole.

Input esclusi.
I campi esclusi dall'analisi e il motivo dell'esclusione per ogni campo.

Criterio per Tabelle di regole, Cloud di termini e Mappe

Fiducia.
La percentuale di previsioni regola corrette.

Supporto alla regola.
La percentuale di casi per cui la regola è vera. Ad esempio, se una regola è "If x=1 and y=2, then z=3," il supporto regola è la percentuale effettiva di casi nei dati per cui x=1, y=2 e z=3.

Guadagno cumulativo.
Il guadagno cumulativo è una misura di quanto la regola migliora la previsione rispetto alla possibilità casuale. È il rapporto delle previsioni corrette rispetto alla occorrenza generale del valore previsto. Il valore deve essere maggiore di 1. Ad esempio, se il valore previsto si verifica il 20% delle volte e la confidenza nella previsione è 80%, il valore del guadagno cumulativo è 4.

Supporto condizione.
La percentuale di casi per cui la condizione esiste. Ad esempio, se una regola è "If x=1 and y=2, then z=3," il supporto condizione è la proporzione di casi nei dati per cui x=1 e y=2.

Distribuibilità.
La percentuale di previsioni non corrette quando sono rispettate le condizioni. La distribuibilità è uguale a (1-confidenza) moltiplicato per il supporto condizione o il supporto condizione meno il supporto regola.

Salva

Salva dati di contesto e mappa come una specifica di mappa
Salva le specifiche di mappa in un file esterno (.mplan). È possibile caricare questo file di specifica di mappa nella procedura guidata per analisi successive. È anche possibile utilizzare il file di specifica della mappa con il comando `SPATIAL ASSOCIATION RULES`.

Copia qualsiasi file di dati e mappa nella specifica
I dati provenienti dai file forma della mappa, dai file di dati esterni e dai dataset utilizzati nella specifica di mappa vengono salvati nel file di specifica della mappa.

Calcolo del punteggio
Salva i valori delle regole migliori, i valori di confidenza per le regole e i valori ID numerici per le regole come nuovi campi nell'origine dati specificata.

Origine dati di cui calcolare il punteggio
L'origine o le origini dati in cui vengono creati i nuovi campi. Se l'origine dati non è aperta nella sessione corrente, viene aperta. È necessario salvare esplicitamente il file modificato per poter salvare i nuovi campi.

Valori obiettivo
Creare i nuovi campi per i campi obiettivo (previsione) selezionati.

- Vengono creati due nuovi campi per ciascun obiettivo: valore previsto e valore di confidenza.
- Per i campi obiettivo continuo (scala), il valore previsto è una stringa che descrive un intervallo di valori. Un valore nel formato "(valore1, valore2]" indica "maggiore di valore1 e minore di o uguale a valore2."

Numero di regole migliori
Creare nuovi campi per il numero di regole migliori specificato. Vengono creati tre nuovi campi per ciascuna regola: valore regola, valore di confidenza e valore ID numerico per la regola.

Prefisso nome
Il prefisso da utilizzare per i nuovi nomi campo.

Creazione regola

I parametri di creazione regola impostano i criteri per le regole di associazione generate.

Elementi per regola
Il numero di valori campo che possono essere inclusi nelle previsioni e condizioni della regola. Il numero totale di elementi non può superare 10. Ad esempio, nella regola "If x=1 and y=2, then z=3", ci sono due elementi condizione e un elemento previsione.

Numero massimo di previsioni.
Il numero massimo di valori campo che possono presentarsi nelle previsioni per una regola.

Numero massimo di condizioni.
Il numero massimo di valori campo che possono presentarsi nelle condizioni per una regola.

Escludi coppia
Esclude le coppie di campi specificate dall'inclusione nella stessa regola.

Criteri della regola

Fiducia.
La confidenza minima che una regola deve avere per essere inclusa nell'output. La confidenza è la percentuale di previsioni corrette.

Supporto alla regola.
Il supporto regola minimo che una regola deve avere per essere inclusa nell'output. Il valore rappresenta la percentuale di casi per cui la regola è vera nei dati osservati. Ad esempio, se una regola è "If x=1 and y=2, then z=3," il supporto regola è la percentuale effettiva di casi nei dati per cui x=1, y=2 e z=3.

Supporto condizione.
Il supporto condizione minimo che una regola deve avere per essere inclusa nell'output. Il valore rappresenta la percentuale di casi per cui la condizione esiste. Ad esempio, se una regola è "If x=1 and y=2, then z=3," il supporto condizione è la percentuale di casi nei dati per cui x=1 e y=2.

Guadagno cumulativo.
Il guadagno cumulativo minimo che una regola deve avere per essere inclusa nell'output. Il guadagno cumulativo è una misura di quanto la regola migliora la previsione nella possibilità casuale. È il rapporto delle previsioni corrette rispetto alla occorrenza generale del valore previsto. Ad esempio, se il valore previsto si verifica il 20% delle volte e la confidenza nella previsione è 80%, il valore del guadagno cumulativo è 4.

Considera come uguale
Identifica le coppie di campi che devono essere trattati come lo stesso campo.

Raccolta e aggregazione

- L'aggregazione è necessaria quando sono presenti più record nei dati rispetto alle funzioni nella mappa. Ad esempio, si dispone di record di dati per singole regioni ma si dispone di una mappa degli stati.
- È possibile specificare il metodo della misura di riepilogo dell'aggregazione per campi ordinali e continui. I campi nominali vengono aggregati in base al valore modale.

Continuo
Per i campi continui (scala), la misura di riepilogo può essere media, mediana o somma.

Ordinale
Per i campi ordinali, la misura di riepilogo può essere, mediana, moda, più alta o più bassa.

Numero di bin
Imposta il numero massimo di bin per i campi continui (scala). I campi continui sono sempre raggruppati o "raccolti" in intervalli di valori. Ad esempio: inferiore o uguale a 5, superiore a 5 e inferiore o uguale a 10, o superiore a 10.

Aggrega la mappa
Applicare l'aggregazione sia ai dati che alle mappe.

Impostazioni personalizzate per campi specifici
È possibile sovrascrivere la misura di riepilogo predefinita e il numero di bin per campi specifici.

- Fare clic sull'icona per aprire la finestra di dialogo **Selettore di campo** e selezionare un campo da aggiungere all'elenco.
- Nella colonna **Aggregazione**, selezionare una misura di riepilogo.
- Per i campi continui, fare clic sul pulsante nella colonna **Bin** per specificare un numero personalizzato di bin per il campo nella finestra di dialogo **Bin**.

Previsione temporale spaziale

Per la previsione temporale spaziale, dopo la definizione delle mappe e delle origini dati occorre effettuare le seguenti operazioni nella procedura guidata:

- Specificare il campo obiettivo, i campi temporali e i predittori facoltativi.
- Definire gli intervalli di tempo o i periodi ciclici per i campi temporali.

In alternativa, è possibile anche:

- Selezionare opzioni di output differenti.
- Personalizzare i parametri di creazione del modello.
- Personalizzare le impostazioni di aggregazione.
- Salvare i valori previsti in un dataset nella sessione corrente o in un file di dati in formato IBM SPSS Statistica .

Selezione campi

L'elenco di campi disponibili include i campi dalle origini dati selezionate. È possibile controllare l'elenco dei campi visualizzati selezionando un'origine dati dall'elenco **Origini dati**.

Destinazione
Un campo obiettivo è obbligatorio. L'obiettivo è il campo per cui vengono previsti i valori.

- Il campo obiettivo deve essere un campo numerico continuo (scala).
- Se sono presenti due origini dati, l'obiettivo corrisponde alle stime di densità kernel e la "Densità" viene visualizzata come il nome dell'obiettivo. Non è possibile modificare questa selezione.

Predittori
È possibile specificare uno o più campi predittore. Questa impostazione è facoltativa.

Campi tempo
È necessario selezionare uno o più campi che rappresentano periodi di tempo o selezionare **Periodi ciclici**.

- Se sono presenti due origini dati, è necessario selezionare i campi temporali da entrambe le origini. Entrambi i campi devono rappresentare lo stesso intervallo.
- Per i periodi ciclici è necessario specificare i campi che definiscono cicli di periodicità nel pannello **Intervalli di tempo** della procedura guidata.

Intervalli di tempo

Le opzioni in questo pannello si basano sulla scelta effettuata tra **Campi tempo** o **Periodo ciclico** nella fase di selezione dei campi.

Campi tempo

Campi tempo selezionati. Se si selezionano dei campi temporali nella fase di selezione dei campi, verranno visualizzati nell'elenco.

Intervallo di tempo. Selezionare l'intervallo di tempo appropriato dall'elenco. A seconda dell'intervallo di tempo è possibile anche specificare altre impostazioni, ad esempio l'intervallo tra le osservazioni (incremento) e il valore iniziale. Questo intervallo di tempo viene utilizzato per tutti i campi temporali selezionati.

- La procedura presume che tutti i casi (record) rappresentano intervalli spaziati equamente.
- In base all'intervallo di tempo selezionato, la procedura può rilevare le osservazioni mancanti o multiple nello stesso intervallo, che necessitano di essere aggregate. Ad esempio, se l'intervallo di tempo è giorni e la data 2014-10-27 è seguita da 2014-10-29, esiste un'osservazione mancante per 2014-10-28. Se l'intervallo di tempo è il mese, più date nello stesso mese vengono aggregate insieme.
- Per alcuni intervalli di tempo, l'impostazione aggiuntiva può definire interruzioni nei normali intervalli spaziati equamente. Ad esempio, se l'intervallo di tempo è giorni, ma sono validi solo i giorni lavorativi, è possibile specificare che ci sono cinque giorni in una settimana e che la settimana comincia il lunedì.
- Se il campo temporale specificato non ha un formato data o ora, l'intervallo di tempo viene automaticamente impostato su **Periodi** e non può essere modificato.

Campi ciclo

Se si seleziona **Periodo ciclico** nella fase di selezione dei campi, è necessario specificare i campi che definiscono i periodi ciclici. Un periodo ciclico identifica la variazione ciclica ripetitiva, ad esempio il numero di mesi dell'anno o il numero di giorni della settimana.

- È possibile specificare fino a tre campi che definiscono i periodi ciclici.
- Il primo campo ciclo rappresenta il livello più alto del ciclo. Ad esempio, se è presente una variazione ciclica per anno, trimestre e mese, il campo che rappresenta l'anno è il primo campo ciclo.
- La lunghezza del ciclo per il primo e il secondo campo ciclo è la periodicità al livello successivo. Ad esempio, se i campi ciclo sono anno, trimestre e mese, la lunghezza del primo ciclo è 4 e del secondo è 3.
- Il valore iniziale per il secondo e terzo campo ciclo è il primo valore in ciascuno di tali periodi ciclici.
- Il valore iniziale e la lunghezza del ciclo devono essere interi positivi.

Aggregazione

- Se si selezionano i **Predittori** nella fase di selezione dei campi, è possibile scegliere per essi il metodo di riepilogo dell'aggregazione.
- L'aggregazione è necessaria quando sono presenti più record in un intervallo di tempo definito. Ad esempio, se l'intervallo di tempo è il mese, più date nello stesso mese vengono aggregate insieme.
- È possibile specificare il metodo della misura di riepilogo dell'aggregazione per campi ordinali e continui. I campi nominali vengono aggregati in base al valore modale.

Continuo
Per i campi continui (scala), la misura di riepilogo può essere media, mediana o somma.

Ordinale
Per i campi ordinali, la misura di riepilogo può essere, mediana, moda, più alta o più bassa.

Impostazioni personalizzate per campi specifici
È possibile sovrascrivere la misura di riepilogo dell'aggregazione predefinita per predittori specifici.

- Fare clic sull'icona per aprire la finestra di dialogo **Selettore di campo** e selezionare un campo da aggiungere all'elenco.
- Nella colonna **Aggregazione**, selezionare una misura di riepilogo.

Output

Mappe

Valori obiettivo.
Mappa dei valori per il campo obiettivo selezionato.

Correlazione
Mappa delle correlazioni.

Cluster
Mappa che evidenzia i cluster di ubicazioni simili tra di loro. Le mappe di cluster sono disponibili solo per i modelli empirici.

Soglia similarità ubicazione.
La similarità richiede la creazione di cluster. Il valore deve essere un numero maggiore di zero e minore di 1.

Specificare il numero massimo di cluster.
Il numero massimo di cluster da visualizzare.

Tabelle di valutazione del modello

Specifiche del modello.
Riepilogo delle specifiche utilizzate per eseguire l'analisi, incluso i campi obiettivo, input e ubicazione.

Riepilogo informazioni temporali.
Identifica i campi temporali e gli intervalli di tempo utilizzati nel modello.

Test degli effetti nella struttura media.
L'output include il valore delle statistiche del test, i gradi di libertà e il livello di significatività per il modello e ciascun effetto.

Struttura media dei coefficienti di modello.
L'output include il valore del coefficiente, l'errore standard, il valore delle statistiche del test, il livello di significatività e gli intervalli di confidenza per ciascun termine del modello.

Coefficienti autoregressivi.
L'output include il valore del coefficiente, l'errore standard, il valore delle statistiche del test, il livello di significatività e gli intervalli di confidenza per ciascun ritardo.

Test di covarianza spaziale.
Per i modelli parametrici basati sul variogramma, visualizza i risultati del test della bontà di adattamento per la struttura di covarianza spaziale. I risultati del test possono determinare se modellare parametricamente la struttura di covarianza spaziale o utilizzare un modello non parametrico.

Covarianza spaziale parametrica.
Per i modelli parametrici basati sul variogramma, visualizza le stime dei parametri per la covarianza spaziale parametrica.

Opzioni modello

Impostazioni modello

Includi automaticamente un'intercetta
Includere l'intercetta nel modello.

Ritardo massimo autoregressione
Il ritardo massimo di autoregressione. Il valore deve essere un intero compreso tra 1 e 5.

Covarianza spaziale
Specifica il metodo di stima per la covarianza spaziale.

Parametrico
Il metodo di stima è parametrico. Il metodo può essere **Gaussiano**, **Esponenziale** o **Esponenziale di potenza**. Per Esponenziale di potenza, è possibile specificare il valore di **Potenza**.

Nonparametrico
Il metodo di stima è non parametrico.

Salva

Salva dati di contesto e mappa come una specifica di mappa
Salva le specifiche di mappa in un file esterno (.mplan). È possibile caricare questo file di specifica di mappa nella procedura guidata per analisi successive. È anche possibile utilizzare il file di specifica della mappa con il comando `SPATIAL TEMPORAL PREDICTION`.

Copia qualsiasi file di dati e mappa nella specifica
I dati provenienti dai file forma della mappa, dai file di dati esterni e dai dataset utilizzati nella specifica di mappa vengono salvati nel file di specifica della mappa.

Calcolo del punteggio
Salva i valori previsti, la varianza e i limiti di confidenza superiore e inferiore per il campo obiettivo nel file di dati selezionato.

- È possibile salvare i valori previsti in un dataset aperto nella sessione corrente o in un file di dati in formato IBM SPSS Statistica .
- Il file di dati non può essere un'origine dati utilizzata nel modello.
- Il file di dati deve contenere tutti i campi temporali e i predittori utilizzati nel modello.
- I valori temporali devono essere superiori ai valori temporali utilizzati nel modello.

Avanzate

Numero massimo di casi con valori mancanti (%)
La percentuale massima di casi con valori mancanti.

Livello di significatività
Il livello di significatività per determinare se è appropriato un modello basato sul variogramma. Il valore deve essere superiore a 0 e inferiore a 1. Il valore predefinito è 0.05. Il livello di significatività viene utilizzato nel test della bontà di adattamento per la struttura di covarianza spaziale. La statistica della bontà di adattamento viene utilizzata per determinare se utilizzare o un modello parametrico o non parametrico.

Fattore di incertezza (%)
Il fattore di incertezza è un valore percentuale che rappresenta la crescita di incertezza per previsioni future. I limiti superiore e inferiore dell'incertezza di previsione aumentano della percentuale specificata ad ogni fase nel futuro.

Fine

Nell'ultima fase della procedura guidata Modellazione geospaziale è possibile eseguire il modello o incollare la sintassi di comando generata in una finestra di sintassi. È possibile modificare e salvare la sintassi generata per utilizzi futuri.

Informazioni particolari

Queste informazioni sono state sviluppate per prodotti e servizi offerti negli Stati Uniti. Questo materiale potrebbe essere disponibile da IBM in altre lingue. Tuttavia, può essere necessario possedere una copia del prodotto o versione del prodotto in tale lingua per potervi accedere.

IBM può non offrire i prodotti, i servizi o le funzioni presentati in questo documento in altri paesi. Consultare il proprio rappresentante locale IBM per informazioni sui prodotti ed i servizi attualmente disponibili nella propria zona. Qualsiasi riferimento ad un prodotto, programma o servizio IBM non implica o intende dichiarare che solo quel prodotto, programma o servizio IBM può essere utilizzato. Qualsiasi prodotto funzionalmente equivalente al prodotto, programma o servizio che non violi alcun diritto di proprietà intellettuale IBM può essere utilizzato. È comunque responsabilità dell'utente valutare e verificare la possibilità di utilizzare altri prodotti, programmi o servizi non IBM.

IBM può avere applicazioni di brevetti o brevetti in corso relativi all'argomento descritto in questo documento. La fornitura del presente documento non concede alcuna licenza a tali brevetti. È possibile inviare per iscritto richieste di licenze a:

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
it

Per richieste di licenze relative ad informazioni double-byte (DBCS), contattare il Dipartimento di Proprietà Intellettuale IBM nel proprio paese o inviare richieste per iscritto a:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

IBM (INTERNATIONAL BUSINESS MACHINES CORPORATION) FORNISCE LA PRESENTE PUBBLICAZIONE "NELLO STATO IN CUI SI TROVA" SENZA GARANZIE DI ALCUN TIPO, ESPRESSE O IMPLICITE, IVI INCLUSE, A TITOLO DI ESEMPIO, GARANZIE IMPLICITE DI NON VIOLAZIONE, DI COMMERCIABILITÀ E DI IDONEITÀ PER UNO SCOPO PARTICOLARE. Alcune giurisdizioni non consentono la rinuncia ad alcune garanzie espresse o implicite in determinate transazioni, pertanto, la presente dichiarazione può non essere applicabile.

Queste informazioni potrebbero contenere imprecisioni tecniche o errori tipografici. Le modifiche alle presenti informazioni vengono effettuate periodicamente; tali modifiche saranno incorporate nelle nuove pubblicazioni della pubblicazione. IBM si riserva il diritto di apportare miglioramenti e/o modifiche al prodotto o al programma descritto nel manuale in qualsiasi momento e senza preavviso.

I riferimenti in queste informazioni a siti Web non IBM vengono forniti solo per comodità e non implicano in alcun modo l'approvazione di tali siti web. I materiali presenti in tali siti web non sono parte dei materiali per questo prodotto IBM e l'utilizzo di tali siti è a proprio rischio.

IBM può utilizzare o distribuire qualsiasi informazione fornita in qualsiasi modo ritenga appropriato senza incorrere in alcun obbligo verso l'utente.

Coloro che detengano la licenza su questo programma e desiderano avere informazioni su di esso allo scopo di consentire: (i) uno scambio di informazioni tra programmi indipendenti ed altri (compreso questo) e (ii) l'utilizzo reciproco di tali informazioni, dovrebbe rivolgersi a:

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119

Armonk, NY 10504-1785
it

Tali informazioni possono essere disponibili, in base ad appropriate clausole e condizioni, includendo in alcuni casi, il pagamento di una tassa.

Il programma concesso in licenza descritto nel presente documento e tutto il materiale concesso in licenza disponibile sono forniti da IBM in base alle clausole dell'Accordo per Clienti IBM (IBM Customer Agreement), dell'IBM IPLA (IBM International Program License Agreement) o qualsiasi altro accordo equivalente tra le parti.

Gli esempi client e dati delle prestazioni citati sono presentati esclusivamente a scopo illustrativo. Gli effettivi risultati delle prestazioni possono variare in base alle configurazioni e alle condizioni operative specifiche.

Le informazioni relative a prodotti non IBM sono ottenute dai fornitori di quei prodotti, dagli annunci pubblicati i da altre fonti disponibili al pubblico. IBM non ha testato quei prodotti e non può confermarne la precisione della prestazione, la compatibilità o qualsiasi altro reclamo relativo ai prodotti non IBM. Le domande sulle funzionalità dei prodotti non IBM devono essere indirizzate ai fornitori di tali prodotti.

Le dichiarazioni relative all'orientamento o alle intenzioni future di IBM sono soggette a modifica o a ritiro senza preavviso e rappresentano solo mete e obiettivi.

Questa pubblicazione contiene esempi di dati e prospetti utilizzati quotidianamente nelle operazioni aziendali. Pertanto, per maggiore completezza, gli esempi includono nomi di persone, società, marchi e prodotti. Tutti i nomi contenuti nel manuale sono fittizi e ogni riferimento a persone o aziende reali è puramente casuale.

LICENZA DI COPYRIGHT:

Queste informazioni contengono programmi applicativi di esempio in linguaggio sorgente, che illustrano tecniche di programmazione su varie piattaforme operative. È possibile copiare, modificare e distribuire questi programmi di esempio sotto qualsiasi forma senza alcun pagamento a IBM, allo scopo di sviluppare, utilizzare, commercializzare o distribuire i programmi applicativi in conformità alle API (application programming interface) a seconda della piattaforma operativa per cui i programmi di esempio sono stati scritti. Questi esempi non sono stati testati approfonditamente tenendo conto di tutte le condizioni possibili. IBM, quindi, non può garantire o sottintendere l'affidabilità, l'utilità o il funzionamento di questi programmi. I programmi di esempio sono forniti "NELLO STATO IN CUI SI TROVANO", senza alcun tipo di garanzia. IBM non intende essere responsabile per alcun danno derivante dall'uso dei programmi di esempio.

Ogni copia o qualsiasi parte di questi programmi di esempio o qualsiasi lavoro derivato, devono contenere le seguenti informazioni relative alle leggi sul diritto d'autore:

© Copyright IBM Corp. 2021. Le porzioni di questo codice derivano da IBM Corp. Programmi di esempio.

© Copyright IBM Corp. 1989 - 2021. Tutti i diritti riservati.

Marchi

IBM, il logo IBM e ibm.com sono marchi o marchi registrati di International Business Machines Corp., registrati in molte giurisdizioni in tutto il mondo. Altri nomi di prodotti e servizi possono essere marchi di IBM o di altre società. Un elenco corrente dei marchi IBM è disponibile sul web in "Copyright and trademark information" all'indirizzo www.ibm.com/legal/copytrade.shtml.

Adobe, il logo Adobe, PostScript e il logo PostScript sono marchi o marchi registrati di Adobe Systems Incorporated negli Stati Uniti e/o in altri paesi.

Intel, il logo Intel, Intel Inside, il logo Intel Inside, Intel Centrino, il logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium sono marchi o marchi registrati di Intel Corporation o delle sue consociate negli Stati Uniti e in altri paesi.

Linux è un marchio registrato di Linus Torvalds negli Stati Uniti e/o negli altri paesi.

Microsoft, Windows, Windows NT e il logo Windows sono marchi di Microsoft Corporation negli Stati Uniti e/o negli altri paesi.

UNIX è un marchio della The Open Group negli Stati Uniti e/o negli altri paesi.

Java e tutti i marchi e i logo basati su Java sono marchi o marchi registrati di Oracle e/o associate.

Indice analitico

A

Accoppiato - Samples *t*- Test 117
Accordo Interrater 264
adattamento automatico della distribuzione
    nella simulazione 294
adattamento della distribuzione
    nella simulazione 294
affidabilità di Spearman-Brown
    in analisi di affidabilità 264
affidabilità di suddivisione a metà
    in analisi di affidabilità 263, 264
Agresti-Caffo
    in Independent - Proporzioni Campioni 113
Agresti-Coull
    in One - Esempio Proporzioni 108
Agresti-Min
    in Pandate - Proporzioni Campioni 110
alfa di Cronbach
    in analisi di affidabilità 263, 264
allocazione della memoria
    nell'analisi cluster TwoStep 222
analisi a risposta multipla
    Risposte multiple: Frequenze 255
    Risposte multiple: Tabelle di contingenza 256
    tabella delle frequenze 255
    tavola di contingenza 256
analisi cluster
    Analisi cluster gerarchica 228
    Analisi del cluster delle k-medie 230
    efficienza 231
Analisi cluster gerarchica
    cluster di appartenenza 230
    dendrogrammi 230
    esempio 228
    funzioni aggiuntive del comando 230
    grafici a ghiacciolo 230
    matrici delle distanze 230
    metodi di raggruppamento 229
    misure della distanza 229
    misure di similarità 229
    orientamento del grafico 230
    pianificazioni di agglomerazione 230
    raggruppamento dei casi in cluster 228
    salvataggio di nuove variabili 230
    Statistiche 228, 230
    trasformazione di misure 229
    trasformazione di valori 229
    variabili 228
Analisi cluster TwoStep
    opzioni 222
    salvataggio in un file di lavoro 223
    salvataggio in un file esterno 223
    Statistiche 223
Analisi del cluster delle k-medie
    cenni generali 230
    cluster di appartenenza 232
Analisi del cluster delle k-medie *(Continua)*
    criteri di convergenza 231
    distanze tra cluster 232
    efficienza 231
    esempi 230
    funzioni aggiuntive del comando 232
    iterazioni 231
    metodi 230
    salvataggio di informazioni del cluster 232
    Statistiche 230, 232
    valori mancanti 232
Analisi del vicino più vicino
    opzioni 209
    output 208
    partizioni 207
    salvataggio di variabili 208
    selezione delle funzioni 207
    vicini 206
    vista modello 209
analisi della varianza
    in ANOVA a una via 120
    in medie 102
    in regressione lineare 172
    in stima di curve 200
analisi delle componenti principali 216, 217
Analisi di affidabilità
    coefficiente di correlazione intraclasse 264
    correlazione e covarianze interelemento 264
    descrittive 264
    esempio 263
    funzioni aggiuntive del comando 267
    Kuder-Richardson 20 264
    Statistiche 263, 264
    T 2 di Hotelling 264
    Tabella ANOVA 264
    Test di additività di Tukey 264
Analisi di potenza
    Statistiche 1
analisi di sensibilità
    nella simulazione 297
analisi di simulazione
    nella simulazione 297
Analisi discriminante
    coefficienti di funzione 213
    criteri 214
    definizione di intervalli 213
    Distanza di Mahalanobis 214
    esempio 212
    esportazione di informazioni dei modelli 215
    funzioni aggiuntive del comando 216
    grafici 215
    Lambda di Wilks 214
    matrice di covarianza 215
    matrici 213
    metodi discriminanti 214
    metodi stepwise 212
    opzioni di visualizzazione 214, 215

Analisi discriminante *(Continua)*
probabilità a priori 215
salvataggio di variabili di classificazione 215
selezione di casi 213
statistica descrittiva 213
Statistiche 212, 213
V di Rao 214
valori mancanti 215
variabili di raggruppamento 212
variabili indipendenti 212
Analisi fattoriale
cenni generali 216
convergenza 217, 218
descrittive 217
esempio 216
formato di visualizzazione dei coefficienti 219
funzioni aggiuntive del comando 219
grafici del caricamento 218
metodi di estrazione 217
metodi di rotazione 218
punteggi fattoriali 219
selezione di casi 217
Statistiche 216, 217
valori mancanti 219
Analisi PROXMAP 141, 143
Analisi ROC
statistiche e grafici 285, 286
analisi serie storiche
previsione 202
previsione di casi 202
ANOVA
in ANOVA a una via 120
in GLM univariato 124
in medie 102
modello 126
nei modelli lineari 166
ANOVA a una via
confronti multipli 121
contrasti 121
contrasti polinomiali 121
funzioni aggiuntive del comando 124
opzioni 123
Statistiche 123
test Post Hoc 121
valori mancanti 123
variabili fattore 120
Anscombe
in One - Esempio Proporzioni 108
asimmetria
in cubi OLAP 104
in descrittive 89
in Esplora 92
in frequenze 87
in medie 102
in Report: Riepiloghi per colonne 261
in Report: Riepiloghi per righe 259
in Riassumi 100
associazione lineare
in tabelle di contingenza 96
autovalori
in analisi fattoriale 217
in regressione lineare 172

B

bagging
nei modelli lineari 162
Binario meta - analisi
Dialogo di analisi 58
Dialogo di contrasto 60
Dialogo di criteri 57
Dialogo di inferenza 59
Dialogo di stampa 63
Finestra di dialogo 64
Finestra di dialogo Bias 60
Finestra di dialogo Trim - and - Fill 61
Finestra di salvataggio 63
Scheda Bubble Plot 79
Scheda Cumulativo Forest Plot 78
Scheda Galbraith Plot 82
Scheda Grafico di Funnel 81
Scheda L'Abb'e Plot 82
Scheda Plot Plot 77
Binomiale adattato Mid-p
in Pandate - Proporzioni Campioni 111
Binomiale esatto
in Pandate - Proporzioni Campioni 111
Bland-Altman
analisi dell'accordo 136, 137
Bonett-Price
in Pandate - Proporzioni Campioni 110
Bonferroni
in ANOVA a una via 121
in GLM 130
bontà di adattamento
in Regressione ordinale 175
boosting
nei modelli lineari 162
Brown-Li-Jeffreys
in Independent - Proporzioni Campioni 113
Builder di simulazioni 291

C

C di Dunnett
in ANOVA a una via 121
in GLM 130
campione di addestramento
nell'analisi del vicino più prossimo 207
campione holdout
nell'analisi del vicino più prossimo 207
campioni dipendenti 250, 253
categoria di riferimento
in GLM 127, 128
chi-quadrato
associazione lineare 96
correzione di continuità di Yates 96
in tabelle di contingenza 96
intervallo atteso. 244
opzioni 244
Pearson 96
per l'indipendenza 96
rapporto di verosimiglianza 96
Statistiche 244
test esatto di Fisher 96
test per un campione 243
valori attesi 244

chi-quadrato *(Continua)*
valori mancanti 244
Chi-quadrato
Test di ipotesi statistica 137, 138
chi-quadrato del rapporto di verosimiglianza
in Regressione ordinale 175
in tabelle di contingenza 96
Chi-quadrato di Pearson
in Regressione ordinale 175
in tabelle di contingenza 96
ChiSquare 137, 138
classificazione
in Analisi ROC 284
nella Curva ROC 287
Clopper-Pearson (esatto)
in One - Esempio Proporzioni 108
clustering
scelta di una procedura 220
visualizzazione di cluster 224
visualizzazione globale 224
coefficiente alfa
in analisi di affidabilità 263, 264
coefficiente di concordanza di Kendall (W)
Test non parametrici di campioni correlati 241
coefficiente di contingenza
in tabelle di contingenza 96
coefficiente di correlazione dei ranghi
in correlazioni bivariate 138
coefficiente di correlazione di Spearman
in correlazioni bivariate 138
in tabelle di contingenza 96
coefficiente di correlazione intraclasse
in analisi di affidabilità 264
coefficiente di correlazione r
in correlazioni bivariate 138
in tabelle di contingenza 96
coefficiente di dispersione (COD)
in Statistica dei rapporti 277
coefficiente di incertezza
in tabelle di contingenza 96
coefficiente di rischio
in tabelle di contingenza 96
coefficiente di variazione (COV)
in Statistica dei rapporti 277
coefficienti beta
in regressione lineare 172
coefficienti di regressione.
in regressione lineare 172
collegamento
in Regressione ordinale 174
colonna totale
nei report 261
confronti multipli
in ANOVA a una via 121
confronti multipli post hoc 121
confronto di gruppi
in cubi OLAP 106
confronto di variabili
in cubi OLAP 106
confronto tra i metodi
nel test di Bland-Altman
136
conteggio atteso
in tabelle di contingenza 97
conteggio osservato
in tabelle di contingenza 97
contrasti
in ANOVA a una via 121
in GLM 127, 128
contrasti di deviazione
in GLM 127, 128
contrasti di differenza
in GLM 127, 128
contrasti di Helmert
in GLM 127, 128
contrasti polinomiali
in ANOVA a una via 121
in GLM 127, 128
contrasti ripetuti
in GLM 127, 128
contrasti semplici
in GLM 127, 128
controllo pagina
nei report di riepilogo per colonne 262
nei report di riepilogo per righe 260
convergenza
in analisi fattoriale 217, 218
nell'Analisi del cluster delle k-medie 231
Correlazione della distanza 159
Correlazione di Pearson
in correlazioni bivariate 138
in tabelle di contingenza 96
correlazioni
di ordine zero 157
in correlazioni bivariate 138
in Correlazioni parziali 156
in tabelle di contingenza 96
nella simulazione 297
Correlazioni di bivariate
coefficienti di correlazione 138
funzioni aggiuntive del comando 141
intervalli di confidenza 140
intervallo di affidabilità 138
livello di significatività 138
opzioni 140
Statistiche 140
valori mancanti 140
correlazioni di ordine zero
in Correlazioni parziali 157
Correlazioni parziali
correlazioni di ordine zero 157
funzioni aggiuntive del comando 157
in regressione lineare 172
opzioni 157
Statistiche 157
valori mancanti 157
correzione di continuità di Yates
in tabelle di contingenza 96
costruzione di termini 126, 176, 177
criteri
nel test di Bland-Altman
136
criteri di informazioni
nei modelli lineari 163
criteri di selezione
in regressione lineare 172
criterio di informazione di Akaike
nei modelli lineari 163

criterio di prevenzione del sovradattamento
nei modelli lineari 163
cronologia delle iterazioni
in Regressione ordinale 175
Cubi OLAP
Statistiche 104
titoli 106
curtosi
in cubi OLAP 104
in descrittive 89
in Esplora 92
in frequenze 87
in medie 102
in Report: Riepiloghi per colonne 261
in Report: Riepiloghi per righe 259
in Riassumi 100
Curva ROC
statistiche e grafici 288

D

d
in tabelle di contingenza 96
D di Somers
in tabelle di contingenza 96
decomposizione gerarchica 126
Definisci insiemi a risposta multipla
categorie 254
dicotomie 254
etichette degli insiemi 254
nomi degli insiemi 254
dendrogrammi
in Analisi cluster gerarchica 230
Descrittive
funzioni aggiuntive del comando 90
ordine di visualizzazione 89
salvataggio dei punteggi z 89
Statistiche 89
deviazione media assoluta (AAD)
in Statistica dei rapporti 277
deviazione standard
in cubi OLAP 104
in descrittive 89
in Esplora 92
in frequenze 87
in medie 102
in Report: Riepiloghi per colonne 261
in Report: Riepiloghi per righe 259
in Riassumi 100
in Statistica dei rapporti 277
DfBeta
in regressione lineare 170
DfFit
in regressione lineare 170
di sicurezza
colonne totale 261
confronto di colonne 261
divisione di valori di colonne 261
moltiplicazione di valori di colonne 261
report di riepilogo per colonne 260
report di riepilogo per righe 258
totali composti 261
Dialogo di analisi
Binario meta - analisi 58
Dialogo di analisi *(Continua)*
Meta - Analisi Continua Effetto Dimensione 49
Meta - Analisi Continuo 42
Meta - Analisi Effetto Binario Dimensione 66
Dialogo di contrasto
Binario meta - analisi 60
Meta - Analisi Continua Effetto Dimensione 51
Meta - Analisi Continuo 43
Meta - Analisi Effetto Binario Dimensione 68
Dialogo di criteri
Binario meta - analisi 57
Meta - Analisi Continuo 41, 49
Meta - Analisi Effetto Binario Dimensione 65
Regressione meta - analisi 74
Dialogo di inferenza
Binario meta - analisi 59
Meta - Analisi Continua Effetto Dimensione 50
Meta - Analisi Continuo 42
Meta - Analisi Effetto Binario Dimensione 67
Regressione meta - analisi 75
Dialogo di stampa
Binario meta - analisi 63
Meta - Analisi Continua Effetto Dimensione 54
Meta - Analisi Continuo 46
Meta - Analisi Effetto Binario Dimensione 71
Regressione meta - analisi 75
differenza meno significativa (LSD)
in ANOVA a una via 121
in GLM 130
differenza meno significativa (LSD) di Fisher
in GLM 130
differenza significativa di Tukey
in ANOVA a una via 121
in GLM 130
differenze tra gruppi
in cubi OLAP 106
differenze tra variabili
in cubi OLAP 106
differenziale di prezzo (PRD)
in Statistica dei rapporti 277
Dimensione degli effetti
in Ponda - Samples T-Test 117
in test T per campioni indipendenti 115
distanza chi-quadrato
in distanze 158
distanza city-block
nell'analisi del vicino più prossimo 206
distanza City-Block
in distanze 158
distanza di Chebychev
in distanze 158
Distanza di Cook
in GLM 133
in regressione lineare 170
Distanza di Mahalanobis
in analisi discriminante 214
in regressione lineare 170
Distanza di Manhattan
nell'analisi del vicino più prossimo 206
Distanza euclidea
in distanze 158
nell'analisi del vicino più prossimo 206
distanza euclidea al quadrato
in distanze 158

distanza Minkowski
in distanze 158
Distanze
calcolo delle distanze tra casi 157
calcolo delle distanze tra variabili 157
esempio 157
funzioni aggiuntive del comando 159
misure di dissimilarità 158
misure di similarità 158
Statistiche 157
trasformazione di misure 158
trasformazione di valori 158
distanze dei vicini più vicini
nell'analisi del vicino più prossimo 211
divisione
divisione tra colonne del report 261
dizionario
Informazioni sui dati 83

E

elenco dei casi 99
eliminazione all'indietro
in regressione lineare 169
equivalenti
nell'analisi del vicino più prossimo 211
errore standard
in Analisi ROC 285, 286
in descrittive 89
in Esplora 92
in frequenze 87
in GLM 133
nella Curva ROC 288
errore standard dell'asimmetria
in cubi OLAP 104
in medie 102
in Riassumi 100
errore standard della curtosi
in cubi OLAP 104
in medie 102
in Riassumi 100
errore standard della media
in cubi OLAP 104
in medie 102
in Riassumi 100
Esplora
funzioni aggiuntive del comando 94
grafici 93
opzioni 94
Statistiche 92
trasformazioni di potenza 93
valori mancanti 94
eta
in medie 102
in tabelle di contingenza 96
eta-quadrato
in medie 102

F

F multiplo di Ryan-Einot-Gabriel-Welsch
in ANOVA a una via 121
in GLM 130
fattore di inflazione della varianza
in regressione lineare 172
fattorizzazione alfa 217
fattorizzazione dell'asse principale 217
fattorizzazione immagine 217
Filtri per serie temporali
Filtro Baxter king 199
Filtro Christiano Fitzgerald 199
Filtro Hodrick Prescott 199
Finestra di dialogo
Binario meta - analisi 64
Meta - Analisi Continua Effetto Dimensione 55
Meta - Analisi Continuo 47
Meta - Analisi Effetto Binario Dimensione 72
Regressione meta - analisi 77
Finestra di dialogo Bias
Binario meta - analisi 60
Meta - Analisi Continua Effetto Dimensione 52
Meta - Analisi Continuo 44
Meta - Analisi Effetto Binario Dimensione 69
Finestra di dialogo Trim - and - Fill
Binario meta - analisi 61
Meta - Analisi Continua Effetto Dimensione 53
Meta - Analisi Continuo 45
Meta - Analisi Effetto Binario Dimensione 70
Finestra di salvataggio
Binario meta - analisi 63
Meta - Analisi Continua Effetto Dimensione 55
Meta - Analisi Continuo 47
Meta - Analisi Effetto Binario Dimensione 72
Regressione meta - analisi 76
Fleiss ' Più Rater Kappa 263, 264
formattazione
colonne nei report 259
Frequenze
formati 89
grafici 88
ordine di visualizzazione 89
soppressione di tabelle 89
Statistiche 87
frequenze attese
in Regressione ordinale 175
frequenze cumulative
in Regressione ordinale 175
frequenze dei cluster
nell'analisi cluster TwoStep 223
frequenze osservate
in Regressione ordinale 175
funzione di densità di probabilità
nella simulazione 299
funzioni di distribuzione cumulativa
nella simulazione 299

G

gamma
in tabelle di contingenza 96
gamma di Goodman e Kruskal
in tabelle di contingenza 96
gamma multipla di Ryan-Einot-Gabriel-Welsch
in ANOVA a una via 121
in GLM 130
gestione del rumore
nell'analisi cluster TwoStep 222

GLM
grafici di profilo 128
modello 126
salvataggio di matrici 133
salvataggio di variabili 133
somma dei quadrati 126
test Post Hoc 130
GLM univariato
contrasti 127, 128
grafici
etichette dei casi 200
Grafici di probabilità 279, 281
in Analisi ROC 284
nella Curva ROC 287
Grafici
Meta - analisi 77
grafici a barre
in frequenze 88
grafici a dispersione
in regressione lineare 170
grafici a ghiacciolo
in Analisi cluster gerarchica 230
grafici a scatole
confronto dei livelli dei fattori 93
confronto di variabili 93
in Esplora 93
nella simulazione 300
grafici a torta
in frequenze 88
grafici del caricamento
in analisi fattoriale 218
grafici delle probabilità normale
in Esplora 93
in regressione lineare 170
P-P 279
Q-Q 281
grafici di diffusione vs. densità
in Esplora 93
grafici di normalità detrendizzati
in Esplora 93
Grafici di probabilità
P-P 279
Q-Q 281
Grafici di probabilità P-P 279
grafici di profilo
in GLM 128
grafici parziali
in regressione lineare 170
grafici ramo-foglia
in Esplora 93
grafici tornado
nella simulazione 300
Grafico a bolle
Regressione meta - analisi 79
grafico a dispersione
nella simulazione 300
Grafico a imbuto 77
grafico dello spazio delle funzioni
nell'analisi del vicino più prossimo 209
Grafico Forest 77
Grafico Forest cumulativo 77
Grafico Galbraith 77

H

H di Kruskal-Wallis
in test per due campioni indipendenti 251
Hauck-Anderson
in Independent - Proporzioni Campioni 113, 114
Hochberg (GT2)
in ANOVA a una via 121
in GLM 130

I

ICC. Consultare il coefficiente di correlazione intraclasse 264
importanza delle variabili
nell'analisi del vicino più prossimo 211
importanza predittore
modelli lineari 166
indice di concentrazione
in Statistica dei rapporti 277
informazioni di diagnostica casewise
in regressione lineare 172
informazioni di diagnostica di collinearità
in regressione lineare 172
Informazioni sui dati
output 84
Statistiche 86
insiemi
nei modelli lineari 164
insiemi a risposta multipla
Informazioni sui dati 83
intervalli del rapporto di verosimiglianza
Test non parametrici a campione singolo 234
Intervalli di Clopper-Pearson
Test non parametrici a campione singolo 234
intervalli di confidenza
in Analisi ROC 285, 286
in ANOVA a una via 123
in correlazioni bivariate 140
in Esplora 92
in GLM 127
in regressione lineare 172
in test T per campioni appaiati 118
in test T per campioni indipendenti 116
in test T per un campione 119
nella Curva ROC 288
salvataggio in regressione lineare 170
Intervalli di Jeffreys
Test non parametrici a campione singolo 234
intervalli di previsione
salvataggio in regressione lineare 170
salvataggio in stima di curve 202
intervallo di affidabilità
in correlazioni bivariate 138
istogrammi
in Esplora 93
in frequenze 88
in regressione lineare 170
iterazioni
in analisi fattoriale 217, 218
nell'Analisi del cluster delle k-medie 231

J

Jeffreys
    in One - Esempio Proporzioni 108

K

kappa
    in tabelle di contingenza 96
kappa di Cohen
    in tabelle di contingenza 96
Kappa pesata
    criteri 268
    esempio 267
    stampare 268
    Statistiche 267, 268
    tavola di contingenza 268
Kappa pesata di Cohen 267, 268
KR20
    in analisi di affidabilità 264
Kuder-Richardson 20 (KR20)
    in analisi di affidabilità 264

L

L'Abb'e Plot 77
lambda
    in tabelle di contingenza 96
lambda di Goodman e Kruskal
    in tabelle di contingenza 96
Lambda di Wilks
    in analisi discriminante 214
Lineare 159
Livello di significatività asintotica 264
Logit
    in One - Esempio Proporzioni 108
Lotti di probabilità Q-Q 281

M

mappa dei quadranti
    nell'analisi del vicino più prossimo 211
MAPPATURA DI PROSSIMITÀ 141, 143
massima verosimiglianza
    in analisi fattoriale 217
massimo
    confronto di colonne del report 261
    in cubi OLAP 104
    in descrittive 89
    in Esplora 92
    in frequenze 87
    in medie 102
    in Riassumi 100
    in Statistica dei rapporti 277
matrice di correlazione
    in analisi discriminante 213
    in analisi fattoriale 216, 217
    in Regressione ordinale 175
matrice di covarianza
    in analisi discriminante 213, 215
    in GLM 133
    in regressione lineare 172
    in Regressione ordinale 175
matrice di modelli
    in analisi fattoriale 216
matrice di trasformazione
    in analisi fattoriale 216
McDonald's Omega
    in analisi di affidabilità 263, 264
McFadden R2
    in Regressione ordinale 175
McNemar
    in Pandate - Proporzioni Campioni 111
McNemar (continuità corretta)
    in Pandate - Proporzioni Campioni 111
media
    di più colonne del report 261
    in ANOVA a una via 123
    in cubi OLAP 104
    in descrittive 89
    in Esplora 92
    in frequenze 87
    in medie 102
    in Report: Riepiloghi per colonne 261
    in Report: Riepiloghi per righe 259
    in Riassumi 100
    in Statistica dei rapporti 277
    sottogruppo 101, 103
media armonica
    in cubi OLAP 104
    in medie 102
    in Riassumi 100
media geometrica
    in cubi OLAP 104
    in medie 102
    in Riassumi 100
media pesata
    in Statistica dei rapporti 277
media ritagliata
    in Esplora 92
mediana
    in cubi OLAP 104
    in Esplora 92
    in frequenze 87
    in medie 102
    in Riassumi 100
    in Statistica dei rapporti 277
mediana dei gruppi
    in cubi OLAP 104
    in medie 102
    in Riassumi 100
Medie
    opzioni 102
    Statistiche 102
medie di gruppi 101, 103
medie di sottogruppi 101, 103
Meta - analisi
    Grafici 77
Meta - Analisi Continua Effetto Dimensione
    Dialogo di analisi 49
    Dialogo di contrasto 51
    Dialogo di inferenza 50
    Dialogo di stampa 54
    Finestra di dialogo 55
    Finestra di dialogo Bias 52
    Finestra di dialogo Trim - and - Fill 53
    Finestra di salvataggio 55

Meta - Analisi Continua Effetto Dimensione *(Continua)*
Scheda Bubble Plot 79
Scheda Cumulativo Forest Plot 78
Scheda Galbraith Plot 82
Scheda Grafico di Funnel 81
Scheda L'Abb'e Plot 82
Scheda Plot Plot 77
Meta - Analisi Continuo
Dialogo di analisi 42
Dialogo di contrasto 43
Dialogo di criteri 41, 49
Dialogo di inferenza 42
Dialogo di stampa 46
Finestra di dialogo 47
Finestra di dialogo Bias 44
Finestra di dialogo Trim - and - Fill 45
Finestra di salvataggio 47
Scheda Bubble Plot 79
Scheda Cumulativo Forest Plot 78
Scheda Galbraith Plot 82
Scheda Grafico di Funnel 81
Scheda L'Abb'e Plot 82
Scheda Plot Plot 77
Meta - Analisi Effetto Binario Dimensione
Dialogo di analisi 66
Dialogo di contrasto 68
Dialogo di criteri 65
Dialogo di inferenza 67
Dialogo di stampa 71
Finestra di dialogo 72
Finestra di dialogo Bias 69
Finestra di dialogo Trim - and - Fill 70
Finestra di salvataggio 72
Scheda Bubble Plot 79
Scheda Cumulativo Forest Plot 78
Scheda Galbraith Plot 82
Scheda Grafico di Funnel 81
Scheda L'Abb'e Plot 82
Scheda Plot Plot 77
Meta-analisi
regressione 39
Risultati binari 39
Risultati continui 39
minimi quadrati generalizzati
in analisi fattoriale 217
minimi quadrati non pesati
in analisi fattoriale 217
minimi quadrati pesati
in regressione lineare 168
minimo
confronto di colonne del report 261
in cubi OLAP 104
in descrittive 89
in Esplora 92
in frequenze 87
in medie 102
in Riassumi 100
in Statistica dei rapporti 277
misura della differenza delle misure
in distanze 158
misura della distanza phi-quadrato
in distanze 158
misura delle differenze dei modelli
in distanze 158
misura di dissimilarità di Lance e Williams
in distanze 158
misure della distanza
in Analisi cluster gerarchica 229
in distanze 158
nell'analisi del vicino più prossimo 206
misure di dispersione
in descrittive 89
in Esplora 92
in frequenze 87
in Statistica dei rapporti 277
misure di distribuzione
in descrittive 89
in frequenze 87
misure di similarità
in Analisi cluster gerarchica 229
in distanze 158
misure di tendenza centrale
in Esplora 92
in frequenze 87
in Statistica dei rapporti 277
modellazione geospaziale 305–316
modellazione spaziale 305
modelli fattoriali completi
in GLM 126
modelli lineari
coefficienti 167
criteri di informazioni 165
importanza predittore 166
insiemi 164
livello di confidenza 163
medie stimate 168
obiettivi 162
opzioni modello 165
preparazione automatica dati 163, 165
previsioni e osservazioni 166
regole di combinazione 164
replica dei risultati 165
residui 166
riepilogo creazione modello 168
riepilogo del modello 165
selezione modello 163
Statistica R-quadrato 165
Tabella ANOVA 166
valori anomali 166
modelli personalizzati
in GLM 126
modello composto
in stima di curve 201
modello cubico
in stima di curve 201
modello di crescita
in stima di curve 201
modello di Guttman
in analisi di affidabilità 263, 264
modello di potenza
in stima di curve 201
modello di regressione ausiliaria
in GLM 135
modello di scala
in Regressione ordinale 176
modello di ubicazione
in Regressione ordinale 176
modello esponenziale

modello esponenziale *(Continua)*
in stima di curve 201
modello inverso
in stima di curve 201
modello lineare
in stima di curve 201
modello logaritmico
in stima di curve 201
modello logistico
in stima di curve 201
modello parallelo
in analisi di affidabilità 263, 264
modello parallelo esatto
in analisi di affidabilità 263, 264
modello quadratico
in stima di curve 201
modello S
in stima di curve 201
modo
in frequenze 87
moltiplicazione
moltiplicazione tra colonne del report 261

N

Newcombe
in Independent - Proporzioni Campioni 113
in Pandate - Proporzioni Campioni 110
Newcombe (continuità corretta)
in Independent - Proporzioni Campioni 113
Newman-Keuls
in GLM 130
non lineare 159
Normalizzazione di Kaiser
in analisi fattoriale 218
numerazione delle pagine
nei report di riepilogo per colonne 262
nei report di riepilogo per righe 260
numero di casi
in cubi OLAP 104
in medie 102
in Riassumi 100

O

One - Proporzioni dei campioni 110

P

percentili
in Esplora 92
in frequenze 87
nella simulazione 300
Percentili 90, 91
percentuali
in tabelle di contingenza 97
percentuali di colonna
in tabelle di contingenza 97
percentuali di riga
in tabelle di contingenza 97
percentuali totali
in tabelle di contingenza 97
phi
in tabelle di contingenza 96
phi *(Continua)*
in tabelle di contingenza 96
PLUM
in Regressione ordinale 173
preparazione automatica dati
nei modelli lineari 165
previsione
in stima di curve 202
primo
in cubi OLAP 104
in medie 102
in Riassumi 100
profondità dell'albero
nell'analisi cluster TwoStep 222
progettazioni per test dell'eteroschedasticità
in GLM 135
Proporzioni campione singolo 106, 108
Proporzioni campioni accoppiati 109
Proporzioni campioni indipendenti 112, 113
Prossimità
in Analisi cluster gerarchica 228
punteggi fattoriali 219
punteggi fattoriali di Anderson-Rubin 219
punteggi fattoriali di Bartlett 219
punteggi z
in descrittive 89
salvataggio come variabili 89
Punteggio 108
Punteggio (continuità corretta) 108
Punteggio Wilson
in One - Esempio Proporzioni 108
Punteggio Wilson (continuità corretta)
in One - Esempio Proporzioni 108

Q

Q di Cochran
in test per diversi campioni dipendenti 253
quartili
in frequenze 87

R

R 2
in medie 102
in regressione lineare 172
modifica R 2 172
R 2 adattato
in regressione lineare 172
R multiplo
in regressione lineare 172
R-E-G-W F
in ANOVA a una via 121
in GLM 130
R-E-G-W Q
in ANOVA a una via 121
in GLM 130
R-quadrato
nei modelli lineari 165
R-quadrato adattato
nei modelli lineari 163
R2 di Nagelkerke
in Regressione ordinale 175

R2 di Snell e Cox
in Regressione ordinale 175
ramificazioni massime
nell'analisi cluster TwoStep 222
rapporto di covarianza
in regressione lineare 170
regole di combinazione
nei modelli lineari 164
regressione
grafici 170
Meta-analisi 39
Regressione lineare 168
regressione multipla 168
Regressione dei minimi quadrati parziali
esportazione di variabili 204
modello 203
Regressione Elastic Net lineare
opzioni 178
partizione 177
variabili 177
Regressione Lasso lineare
opzioni 181
partizione 181
variabili 181
Regressione lineare
blocchi 168
esportazione di informazioni dei modelli 170
funzioni aggiuntive del comando 173
grafici 170
metodi di selezione delle variabili 169, 173
pesi 168
residui 170
salvataggio di nuove variabili 170
Statistiche 172
valori mancanti 173
variabile di selezione 170
Regressione Lineare Ridge
opzioni 184
partizione 184
variabili 184
Regressione meta - analisi
Dialogo di criteri 74
Dialogo di inferenza 75
Dialogo di stampa 75
Finestra di dialogo 77
Finestra di salvataggio 76
Grafico a bolle 79
regressione multipla
in regressione lineare 168
Regressione ordinale
collegamento 174
funzioni aggiuntive del comando 177
modello di scala 176
modello di ubicazione 176
opzioni 174
Statistiche 173
relazione spaziale 141
Report : Riepiloghi per righe
colonne di dati 258
colonne di interruzione 258
controllo pagina 259
formato colonne 259
funzioni aggiuntive del comando 263
layout di pagina 260
Report : Riepiloghi per righe *(Continua)*
numerazione delle pagine 260
ordinamento delle sequenze 258
piè di pagina 260
spaziatura interruzione 259
titoli 260
valori mancanti 260
variabili nei titoli 260
report di riepilogo per colonne 260
Report: Riepiloghi per colonne
colonne totale 261
controllo pagina 262
formato colonne 259
funzioni aggiuntive del comando 263
layout di pagina 260
numerazione delle pagine 262
totale finale 262
totali parziali 262
valori mancanti 262
residui
in tabelle di contingenza 97
salvataggio in regressione lineare 170
salvataggio in stima di curve 202
residui cancellati
in GLM 133
in regressione lineare 170
residui di Pearson
in Regressione ordinale 175
residui non standardizzati
in GLM 133
residui standardizzati
in GLM 133
in regressione lineare 170
residui studentizzati
in regressione lineare 170
rho
in correlazioni bivariate 138
in tabelle di contingenza 96
Ricerca 137, 138
ridimensiona
in analisi di affidabilità 263
in Kappa pesata 267
in scaling multidimensionale 269
Riepiloga
opzioni 99
Statistiche 100
riepilogo degli errori
nell'analisi del vicino più prossimo 212
rischio
in tabelle di contingenza 96
Risposta multipla
funzioni aggiuntive del comando 257
Risposte multiple: Frequenze
valori mancanti 255
Risposte multiple: Tabelle di contingenza
associazione di variabili negli insiemi a risposta multipla 257
definizione dei bin dei valori 257
percentuali basate sui casi 257
percentuali basate sulle risposte 257
percentuali delle celle 257
valori mancanti 257
Risultati binari
dati di dimensione dell'effetto pre - calcolato 39

Risultati binari *(Continua)*
dati di dimensione dell'effetto raw 39
Risultati continui
dati di dimensione dell'effetto pre - calcolato 39
dati di dimensione dell'effetto raw 39
Meta-analisi 39
rotazione equamax
in analisi fattoriale 218
rotazione oblimin diretta
in analisi fattoriale 218
rotazione quartimax
in analisi fattoriale 218
rotazione varimax
in analisi fattoriale 218

S

S-stress
in scaling multidimensionale 269
scala
in cubi OLAP 104
in descrittive 89
in frequenze 87
in medie 102
in Riassumi 100
in Statistica dei rapporti 277
Scaling multidimensionale
condizionalità 271
creazione di matrici delle distanze 270
criteri 271
definizione della forma dei dati 270
dimensioni 271
esempio 269
funzioni aggiuntive del comando 271
livelli di misurazione 271
misure della distanza 270
modelli di scaling 271
opzioni di visualizzazione 271
Statistiche 269
trasformazione di valori 270
Scheda Bubble Plot
Binario meta - analisi 79
Meta - Analisi Continua Effetto Dimensione 79
Meta - Analisi Continuo 79
Meta - Analisi Effetto Binario Dimensione 79
Scheda Cumulativo Forest Plot
Binario meta - analisi 78
Meta - Analisi Continua Effetto Dimensione 78
Meta - Analisi Continuo 78
Meta - Analisi Effetto Binario Dimensione 78
Scheda Galbraith Plot
Binario meta - analisi 82
Meta - Analisi Continua Effetto Dimensione 82
Meta - Analisi Continuo 82
Meta - Analisi Effetto Binario Dimensione 82
Scheda Grafico di Funnel
Binario meta - analisi 81
Meta - Analisi Continua Effetto Dimensione 81
Meta - Analisi Continuo 81
Meta - Analisi Effetto Binario Dimensione 81
Scheda L'Abb'e Plot
Binario meta - analisi 82
Meta - Analisi Continua Effetto Dimensione 82
Meta - Analisi Continuo 82
Scheda L'Abb'e Plot *(Continua)*
Meta - Analisi Effetto Binario Dimensione 82
Scheda Plot Plot
Binario meta - analisi 77
Meta - Analisi Continua Effetto Dimensione 77
Meta - Analisi Continuo 77
Meta - Analisi Effetto Binario Dimensione 77
selezione delle funzioni
nell'analisi del vicino più prossimo 211
selezione in avanti
in regressione lineare 169
nell'analisi del vicino più prossimo 207
selezione k
nell'analisi del vicino più prossimo 211
selezione k e selezione delle funzioni
nell'analisi del vicino più prossimo 212
selezione stepwise
in regressione lineare 169
simulazione
adattamento della distribuzione 294
analisi di sensibilità 297
analisi di simulazione 297
Builder di simulazioni 291
campionamento delle code 298
correlazioni tra gli input 297
creazione di nuovi input 293
creazione di un piano di simulazione 289, 290
criteri di arresto 298
editor di equazioni 292
esecuzione di un piano di simulazione 290, 302
formati di visualizzazione per obiettivi e input 300
funzione di densità di probabilità 299
funzione di distribuzione cumulativa (CDF) 299
grafici a dispersione 300
grafici a scatole 300
grafici interattivi 304
grafici tornado 300
grafici: opzioni 305
impostazione di modelli 291
modelli supportati 291
output 299, 300
percentili delle distribuzioni di obiettivo 300
personalizzazione dell'adattamento della distribuzione 296
riadattamento delle distribuzioni ai nuovi dati 302
risultati dell'adattamento della distribuzione 296
salva dati simulati 301
salva piano di simulazione 301
simulazione Monte Carlo 288
soglia iniziale
nell'analisi cluster TwoStep 222
somma
in cubi OLAP 104
in descrittive 89
in frequenze 87
in medie 102
in Riassumi 100
somma dei quadrati
in GLM 126
sottoinsiemi migliori
nei modelli lineari 163
stampare
nel test di Bland-Altman 137

standardizzazione
nell'analisi cluster TwoStep 222
standardizzazione di valori
in descrittive 89
statistica descrittiva
in descrittive 89
in Esplora 92
in frequenze 87
in Riassumi 100
in Statistica dei rapporti 277
nell'analisi cluster TwoStep 223
statistica di Brown-Forsythe
in ANOVA a una via 123
statistica di Cochran
in tabelle di contingenza 96
statistica di Mantel-Haenszel
in tabelle di contingenza 96
statistica di Welch
in ANOVA a una via 123
statistica F
nei modelli lineari 163
Statistiche dei rapporti
Statistiche 277
statistiche delle proporzioni di colonna
in tabelle di contingenza 97
statistiche Durbin-Watson
in regressione lineare 172
statistiche R
in medie 102
in regressione lineare 172
stepwise in avanti
nei modelli lineari 163
Stima di curve
analisi della varianza 200
inclusione di costanti 200
modelli 201
previsione 202
salvataggio degli intervalli di previsione 202
salvataggio dei residui 202
salvataggio di valori attesi 202
stimatore di Tuckey a doppio peso
in Esplora 92
Stimatore M decrescente di Hampel
in Esplora 92
stimatore M di Andrew
in Esplora 92
stimatore M di Huber
in Esplora 92
Stimatori M
in Esplora 92
stime del parametro
in Regressione ordinale 175
Stime di Hodges-Lehman
Test non parametrici di campioni correlati 241
strati
in tabelle di contingenza 95
stress
in scaling multidimensionale 269
Student-Newman-Keuls
in ANOVA a una via 121
in GLM 130
studio dei casi di controllo
Test T a campioni accoppiati 117
studio di confronti tra coppie
studio di confronti tra coppie *(Continua)*
in test T per campioni appaiati 117
Successioni di Wald-Wolfowitz
in test per due campioni indipendenti 249

T

T 2 di Hotelling
in analisi di affidabilità 263, 264
T3 di Dunnett
in ANOVA a una via 121
in GLM 130
tabella delle frequenze
in Esplora 92
in frequenze 86
tabella di classificazione
nell'analisi del vicino più prossimo 212
tabelle di contingenza 94
Tabelle di contingenza
formati 99
grafici a barre raggruppate 95
soppressione di tabelle 94
Statistiche 96
strati 95
variabili di controllo 95
visualizzazione cella 97
Tamhane (T2)
in ANOVA a una via 121
in GLM 130
tau di Goodman e Kruskal
in tabelle di contingenza 96
tau di Kruskal
in tabelle di contingenza 96
tau-b
in tabelle di contingenza 96
Tau-b di Kendall
in correlazioni bivariate 138
in tabelle di contingenza 96
tau-c
in tabelle di contingenza 96
Tau-c di Kendall
in tabelle di contingenza 96
tavola di contingenza
in tabelle di contingenza 94
risposta multipla 256
termini di interazione 126, 176, 177
test b di Tukey
in ANOVA a una via 121
in GLM 130
test binomiale
Test non parametrici a campione singolo 234
Test binomiale
dicotomie 244
funzioni aggiuntive del comando 245
opzioni 245
Statistiche 245
valori mancanti 245
Test dei ranghi con segno di Wilcoxon
in test per due campioni dipendenti 250
Test non parametrici a campione singolo 234
Test non parametrici di campioni correlati 241
test dei segni
in test per due campioni dipendenti 250
Test non parametrici di campioni correlati 241

test del chi-quadrato
Test non parametrici a campione singolo 234, 235
test della mediana
in test per due campioni indipendenti 251
test delle linee parallele
in Regressione ordinale 175
Test delle reazioni estreme di Moses
in test per due campioni indipendenti 249
test delle successioni
Test non parametrici a campione singolo 234, 235
Test delle successioni
funzioni aggiuntive del comando 246
opzioni 246
punti di divisione 246
Statistiche 246
valori mancanti 246
Test di additività di Tukey
in analisi di affidabilità 263, 264
Test di due campioni indipendenti
definizione dei gruppi 250
funzioni aggiuntive del comando 250
opzioni 250
Statistiche 250
tipi di test 249
valori mancanti 250
variabili di raggruppamento 250
test di intervallo multiplo di Duncan
in ANOVA a una via 121
in GLM 130
Test di Kolmogorov-Smirnov
Test di Lilliefors 235
Test non parametrici a campione singolo 234, 235
Test di Kolmogorov-Smirnov a campione singolo
distribuzione del test 247
funzioni aggiuntive del comando 249
opzioni 248
Statistiche 248
Test di Lilliefors 247, 248
valori mancanti 248
test di Levene
in ANOVA a una via 123
in Esplora 93
test di Lilliefors
in Esplora 93
Test di Lilliefors 235, 247, 248
test di linearità
in medie 102
test di McNemar
in tabelle di contingenza 96
in test per due campioni dipendenti 250
Test non parametrici di campioni correlati 241
test di normalità
in Esplora 93
test di omogeneità della varianza
in ANOVA a una via 123
test di omogeneità marginale
in test per due campioni dipendenti 250
Test non parametrici di campioni correlati 241
test di Scheffé
in ANOVA a una via 121
in GLM 130
test di sfericità di Bartlett
in analisi fattoriale 217
test di Shapiro-Wilk
in Esplora 93
test esatto di Fisher
in tabelle di contingenza 96
test M di Box
in analisi discriminante 213
test non parametrici
chi-quadrato 243
Test delle successioni 246
Test di due campioni indipendenti 249
Test di Kolmogorov-Smirnov a campione singolo 247
Test per diversi campioni correlati 253
Test per diversi campioni indipendenti 251
Test per due campioni correlati 250
Test non parametrici a campione singolo
Campi 233
test binomiale 234
test del chi-quadrato 235
test delle successioni 235
Test di Kolmogorov-Smirnov 235
Test non parametrici di campioni correlati
Campi 240
test di McNemar 241
Test Q di Cochran 242
Test non parametrici di campioni indipendenti
Scheda Campi 238
Test per confronti pairwise di Gabriel
in ANOVA a una via 121
in GLM 130
Test per confronti pairwise di Games e Howell
in ANOVA a una via 121
in GLM 130
Test per diversi campioni correlati
funzioni aggiuntive del comando 254
Statistiche 253
tipi di test 253
Test per diversi campioni indipendenti
definizione dell'intervallo 252
funzioni aggiuntive del comando 253
opzioni 252
Statistiche 252
tipi di test 252
valori mancanti 252
variabili di raggruppamento 252
Test per due campioni correlati
funzioni aggiuntive del comando 251
opzioni 251
Statistiche 251
tipi di test 251
valori mancanti 251
test per l'indipendenza
chi-quadrato 96
Test Q di Cochran
Test non parametrici di campioni correlati 241, 242
test t
in test T per campioni appaiati 117
in test T per campioni indipendenti 115
in test T per un campione 118
Test T a campione singolo
funzioni aggiuntive del comando 118, 119
intervalli di confidenza 119
opzioni 119
valori mancanti 119

Test T a campioni accoppiati
opzioni 118
selezione di variabili appaiate 117
valori mancanti 118
Test T a campioni indipendenti
definizione dei gruppi 116
intervalli di confidenza 116
opzioni 116
valori mancanti 116
variabili di raggruppamento 116
variabili stringa 116
test t di Dunnett
in ANOVA a una via 121
in GLM 130
test t di Sidak
in ANOVA a una via 121
in GLM 130
test t di Student 115
test t di Waller-Duncan
in ANOVA a una via 121
in GLM 130
test t dipendente
in test T per campioni appaiati 117
test t: due campioni
in test T per campioni indipendenti 115
Test Wilcoxon a campione singolo 237
testi di Friedman
in test per diversi campioni dipendenti 253
Test non parametrici di campioni correlati 241
titoli
in cubi OLAP 106
tolleranza
in regressione lineare 172
Tolleranza
Regressione lineare 173
totali finali
nei report di riepilogo per colonne 262
totali parziali
nei report di riepilogo per colonne 262
tracciare
nel test di Bland-Altman
137
TSF 199

U

U di Mann-Whitney
in test per due campioni indipendenti 249
ultimo
in cubi OLAP 104
in medie 102
in Riassumi 100

V

V
in tabelle di contingenza 96
V di Cramér
in tabelle di contingenza 96
V di Rao
in analisi discriminante 214
valori anomali
in Esplora 92
valori anomali *(Continua)*
in regressione lineare 170
nell'analisi cluster TwoStep 222
valori attesi
salvataggio in regressione lineare 170
salvataggio in stima di curve 202
valori attesi pesati
in GLM 133
valori di leva
in GLM 133
in regressione lineare 170
valori estremi
in Esplora 92
valori mancanti
in analisi fattoriale 219
in Analisi ROC 285, 286
in ANOVA a una via 123
in correlazioni bivariate 140
in Correlazioni parziali 157
in Esplora 94
in Independent - Proporzioni Campioni 114
in One - Esempio Proporzioni 109
in Pandate - Proporzioni Campioni 111
in regressione lineare 173
in Report: Riepiloghi per righe 260
in Risposte multiple: Frequenze 255
in Risposte multiple: Tabelle di contingenza 257
in test di Kolmogorov-Smirnov per un campione
248
in test per due campioni dipendenti 251
in test per due campioni indipendenti 250
in test T per campioni appaiati 118
in test T per campioni indipendenti 116
in test T per un campione 119
nei report di riepilogo per colonne 262
nei test per diversi campioni indipendenti 252
nel test binomiale 245
nel test del chi-quadrato 244
nel Test delle successioni 246
nell'analisi del vicino più prossimo 209
nella Curva ROC 288
Valori mancanti
in percentili 91
variabile di selezione
in regressione lineare 170
variabili di controllo
in tabelle di contingenza 95
varianza
in cubi OLAP 104
in descrittive 89
in Esplora 92
in frequenze 87
in medie 102
in Report: Riepiloghi per colonne 261
in Report: Riepiloghi per righe 259
in Riassumi 100
vista modello
nell'analisi del vicino più prossimo 209
visualizzatore cluster
capovolgi cluster e funzioni 225
cenni generali 224
confronto di cluster 227
dimensione dei cluster 226
distribuzione delle celle 227

visualizzatore cluster *(Continua)*
filtraggio dei record 228
importanza predittore 226
importanza predittore nei cluster, visualizzazione 226
informazioni sui modelli di cluster 223
ordina cluster 226
ordina contenuto della cella 226
ordina funzioni 225
ordinamento della visualizzazione dei cluster 226
ordinamento della visualizzazione delle funzioni 225
riepilogo del modello 224
trasponi cluster e funzioni 225
utilizzo 227
vista Centri cluster 225
vista cluster 225
vista Confronto tra cluster 227
vista di base 226
vista Dimensioni cluster 226
vista Distribuzione delle celle 227
vista riepilogo 224
visualizzazione contenuto cella 226
visualizzazione
modelli di clustering 224

W

W di Kendall
in test per diversi campioni dipendenti 253
Wald
in Independent - Proporzioni Campioni 113, 114
in One - Esempio Proporzioni 108
in Pandate - Proporzioni Campioni 110, 111
Wald (continuità corretta)
in Independent - Proporzioni Campioni 113, 114
in One - Esempio Proporzioni 108
in Pandate - Proporzioni Campioni 110, 111
Wald H0
in Independent - Proporzioni Campioni 114
Wald H0(continuità corretta)
in Independent - Proporzioni Campioni 114

X

X^2 137, 138

Z

Z di Kolmogorov-Smirnov
in test di Kolmogorov-Smirnov per un campione 247
in test per due campioni indipendenti 249