

IBM SPSS Categories 25

IBM

Comunicado

Antes de usar estas informações e o produto suportado por elas, leia as informações nos “Avisos” na página 49.

Informações sobre o produto

Esta edição aplica-se à versão 25, liberação 0, modificação 0 do IBM SPSS Statistics e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

Índice

Categorias (Categories) 1

Introdução aos procedimentos de ajuste de escala ideal para dados categóricos	1
O que é o ajuste de escala ideal?	1
Por que usar o ajuste de escala ideal?	1
Nível de ajuste de escala ideal e nível de medição	2
Qual Procedimento É Melhor para o Seu Aplicativo?	5
Proporção de aspecto nos gráficos de ajuste de escala ideal	10
Regressão categórica (CATREG)	10
Definir escala na regressão categórica	11
Discretização de regressão categórica	12
Valores omissos da regressão categórica	12
Opções de regressão categórica	13
Regularização de regressão categórica	13
Saída de regressão categórica	14
Salvamento de regressão categórica	15
Gráficos de transformação de regressão categórica	15
Recursos adicionais do comando CATREG	15
Análise de componentes principais para dados categóricos (CATPCA)	16
Definir escala e ponderação em CATPCA	17
Discretização de análise de componentes principais para dados categóricos	18
Valores omissos da análise de componentes principais para dados categóricos	18
Opções de análise de componentes principais para dados categóricos	19
Saída de análise de componentes principais para dados categóricos	20
Salvamento da análise de componentes principais para dados categóricos	21
Gráficos de objeto de análise de componentes principais para dados categóricos	21
Gráficos de categoria de Análise de componentes principais para dados categóricos	21
Gráficos de carregamento da análise de componentes principais para dados categóricos	22
Bootstrap de análise de componentes principais para dados categóricos	22
Recursos adicionais do comando CATPCA	23
Análise de correlação canônica não linear (GERAIS)	23
Definir intervalo e escala	25
Definir intervalo	25
Opções de análise de correlação canônica não linear	25
Recursos adicionais de comando OVERALS	26
Análise de correspondências	26
Definir intervalo de linha na análise de correspondência	28
Definir intervalo da coluna em análise de correspondência	28

Modelo de análise de correspondência	28
Estatísticas de análise de correspondência	29
Gráficos de análise de correspondência	30
Recursos adicionais do comando CORRESPONDENCE	30
Análise de correspondências múltipla	31
Definir ponderação variável na análise de múltipla correspondência	32
Discretização de análise de correspondências múltipla	32
Valores omissos da Análise de múltipla correspondência	32
Opções de análise de múltipla correspondência	33
Saída de análise de múltipla correspondência	34
Salvamento da análise de múltipla correspondência	34
Gráficos de objetos de múltiplas análises de correspondência	35
Gráficos da variável de análise de múltiplas correspondências	35
Recursos adicionais do comando MULTIPLE CORRESPONDENCE	35
Ajuste de escala multidimensional (PROXSCAL)	35
Proximidades em matrizes em colunas	37
Proximidades em colunas	37
Proximidades em uma coluna	37
Criar proximidades a partir de dados	38
Criar medida a partir de dados	38
Definir um modelo de ajuste de escala multidimensional	38
Restrições de ajuste de escala multidimensional	39
Opções de ajuste de escala multidimensional	40
Gráficos de ajuste de escala multidimensional, versão 1	40
Gráficos de ajuste de escala multidimensional, versão 2	41
Saída de ajuste de escala multidimensional	41
Recursos adicionais de comandos do PROXSCAL	42
Desdobramento multidimensional (PREFSCAL)	42
Definir um modelo de Desdobramento multidimensional	43
Restrições de desdobramento multidimensional	44
Opções de desdobramento multidimensional	44
Gráficos de desdobramento multidimensional	45
Saída de Desdobramento multidimensional	46
Recursos adicionais do comando PREFSCAL	46

Avisos 49

Marcas comerciais	51
-----------------------------	----

Índice Remissivo 53

Categorias (Categories)

Os recursos de categorias a seguir estão incluídos em SPSS Statistics Professional Edition ou a opção Categorias.

Introdução aos procedimentos de ajuste de escala ideal para dados categóricos

Os procedimentos de categoria usam ajuste de escala ideal para analisar dados que são difíceis ou impossíveis de os procedimentos estatísticos padrão analisarem. Este capítulo descreve o que cada procedimento faz, as situações em que cada procedimento é mais apropriado, os relacionamentos entre os procedimentos e os relacionamentos destes procedimentos para suas contrapartes estatísticas padrão.

Nota: Esses procedimentos e sua implementação no IBM® SPSS Statistics foram desenvolvidas pelo Data Theory Scaling System Group (DTSS), consistindo de membros dos departamentos de Education and Psychology, Faculty of Social and Behavioral Sciences, Leiden University.

O que é o ajuste de escala ideal?

A ideia por trás do ajuste de escala ideal é atribuir quantificações numéricas para as categorias de cada variável, permitindo, assim, que os procedimentos padrão sejam usados para obter uma solução sobre as variáveis quantificadas.

Os valores de escala ideais são atribuídos às categorias de cada variável, com base no critério de otimização do procedimento em uso. Ao contrário dos rótulos originais das variáveis nominais ou ordinais na análise, estes valores da escala têm propriedades métricas.

Na maioria dos procedimentos de Categorias, a quantificação ideal para cada variável em escala é obtida por meio de um método iterativo chamado **alternando quadrados mínimos**, no qual, após as quantificações atuais serem usadas para encontrar uma solução, são atualizadas usando essa solução. As quantificações atualizadas são, então, usadas para encontrar uma nova solução, que é usada para atualizar as quantificações, e assim por diante, até que algum critério seja alcançado sinalizando o processo para parar.

Por que usar o ajuste de escala ideal?

Os dados categóricos são frequentemente encontrados em pesquisas de mercado, pesquisas de opinião e pesquisa em ciências sociais e comportamentais. De fato, muitos pesquisadores lidam quase exclusivamente com dados categóricos.

Embora as adaptações da maioria dos modelos padrão existam especificamente para analisar dados categóricos, elas muitas vezes não executam muito bem os conjuntos de dados que apresentam:

- Pouquíssimas observações
- Muitas variáveis
- Muitos valores por variável

Ao quantificar categorias, as técnicas de ajuste de escala ideais evitam problemas nestas situações. Além disso, elas são úteis, mesmo quando técnicas especializadas são apropriadas.

Em vez de interpretar as estimativas paramétrica, a interpretação do resultado de ajuste de escala ideal é muitas vezes baseado em exibições gráficas. As técnicas de ajuste de escala ideal oferecem excelentes análises exploratórias, que complementam outros modelos do IBM SPSS Statistics também. Ao estreitar o

foco de sua investigação, visualizar seus dados por meio de ajuste de escala ideal pode formar a base de uma análise que gira em torno de interpretação dos parâmetros do modelo.

Nível de ajuste de escala ideal e nível de medição

Este pode ser um conceito muito confuso quando são usados primeiro os procedimentos de categoria. Ao especificar o nível, especifique não o nível em que as variáveis são *medidas* mas o nível no qual elas estão *escaladas*. A ideia é que as variáveis a serem quantificadas possam ter relações não lineares, independentemente de como elas são medidas.

Para propósitos de Categorias, há três níveis básicos de medida:

- O nível **nominal** implica que os valores de uma variável representam categorias não ordenadas. Os exemplos de variáveis que podem ser nominais são região, CEP, afiliação religiosa e várias categorias de opção.
- O nível **ordinal** implica que os valores de uma variável representam categorias ordenadas. Os exemplos incluem escalas de atitude que representam o grau de satisfação ou confiança e os escores de classificação de preferência.
- O nível **numérico** implica que os valores de uma variável representam categorias ordenadas com uma métrica significativa, de modo que as comparações de distância entre categorias são apropriadas. Os exemplos incluem idade em anos e receita em milhares de dólares.

Por exemplo, suponha que as variáveis *região*, *cargo* e *idade* sejam codificados conforme mostrado na tabela a seguir.

Tabela 1. Esquema de codificação para região, cargo e idade

Código da região	Valor de região	Código do cargo	Valor do cargo	Idade
1	Norte	1	estagiário	20
2	Sul	2	representante de vendas	22
3	Leste	3	gerenciador	25
4	Oeste			27

Os valores representam as categorias de cada variável. *Região* seria uma variável nominal. Há quatro categorias de *região*, sem nenhuma ordenação intrínseca. Os valores de 1 a 4 simplesmente representam as quatro categorias; o esquema de codificação é completamente arbitrário. *Cargo*, por outro lado, poderia ser assumido para ser uma variável ordinal. As categorias originais formam uma progressão de estagiário a gerente. Os códigos maiores representam um cargo mais alto na escada corporativa. No entanto, somente as informações da ordem são conhecidas, nada pode ser dito sobre a distância entre as categorias adjacentes. Ao contrário, *idade* poderia ser assumido para ser uma variável numérica. No caso de *idade*, as distâncias entre os valores são intrinsecamente significativas. A distância entre 20 e 22 é a mesma distância entre 25 e 27, enquanto que a distância entre 22 e 25 é maior que qualquer uma destas.

Selecionando o nível de ajuste de escala ideal

É importante compreender que não existem propriedades intrínsecas de uma variável que predefinam automaticamente qual nível de ajuste de escala ideal deve ser especificado para ela. É possível explorar seus dados de qualquer forma que faça sentido e fazer uma interpretação mais fácil. Ao analisar uma variável de nível numérico no nível ordinal, por exemplo, o uso de uma transformação não linear pode permitir uma solução em menos dimensões.

Os dois exemplos a seguir ilustram como o nível de medição "óbvio" pode não ser o melhor nível de ajuste de escala ideal. Suponha que uma variável ordene objetos por faixa etária. Embora a idade possa ser escalada como uma variável numérica, pode ser verdade que, para as pessoas com menos de 25 anos, a segurança tem uma relação positiva com a idade, enquanto que, para pessoas com mais de 60 anos, a segurança tem uma relação negativa com a idade. Neste caso, talvez seja melhor tratar a idade como uma variável nominal.

Como outro exemplo, uma variável que ordena as pessoas por preferência política parece ser essencialmente nominal. No entanto, se você ordenar os partidos de esquerda política para a direita política, poderá desejar que a quantificação dos partidos respeite esta ordem, usando um nível ordinal de análise.

Mesmo que não existam propriedades predefinidas de uma variável que a torne exclusivamente um nível ou outro, existem algumas diretrizes gerais para ajudar o usuário iniciante. Com uma quantificação unicamente nominal, você não costuma saber a ordem das categorias, embora seja desejado que a análise imponha uma. Se a ordem das categorias for conhecida, deve-se tentar a quantificação ordinal. Se as categorias forem não ordenáveis, será possível tentar a quantificação de multinomial.

Gráficos de transformação

Os diferentes níveis nos quais cada variável pode ser escalada impõem diferentes restrições nas quantificações. Os gráficos de transformação ilustram o relacionamento entre as quantificações e as categorias originais resultantes do nível de ajuste de escala ideal selecionado. Por exemplo, os resultados do gráfico de transformação linear quando uma variável é tratada como numérica. As variáveis tratadas como ordinais resultam em um gráfico de transformação não decrescente. Os gráficos de transformação para variáveis tratadas nominalmente que são em formato de U (ou o inverso) exibem um relacionamento quadrático. As variáveis nominais também poderiam render gráficos de transformação sem tendências aparentes, alterando a ordem das categorias completamente. A figura a seguir exibe um gráfico de transformação de amostra.

Os gráficos de transformação são particularmente colocados em conjunto para determinar quão bem o nível de ajuste de escala ideal selecionado executa. Se várias categorias receberem quantificações semelhantes, reduzir essas categorias para uma categoria pode ser garantido. Como alternativa, se uma variável tratada como nominal recebe quantificações que exibem uma tendência crescente, uma transformação ordinal poderá resultar em um ajuste semelhante. Se esta tendência for linear, o tratamento numérico poderá ser apropriado. No entanto, se a redução das categorias ou a mudança dos níveis de escala forem garantidos, a análise não mudará significativamente.

Códigos de categoria

Alguns cuidados devem ser tomados ao codificar variáveis categóricas porque alguns esquemas de codificações podem produzir saída indesejada ou análises incompletas. Esquemas de codificação possíveis para *cargo* são exibidos na tabela a seguir.

Tabela 2. Esquemas de codificação alternativos para cargos

Categoria	A	B	C	E
estagiário	1	1	5	1
representante de vendas	2	2	6	5
gerenciador	3	7	7	3

Alguns procedimentos de Categorias requerem que o intervalo de cada variável usada seja definido. Qualquer valor fora deste intervalo é tratado como um valor omissivo. O valor de categoria mínimo é sempre 1. O valor de categoria máximo é fornecido pelo usuário. Este valor não é o *número* de categorias para uma variável—ele é o valor de categoria *maior*. Por exemplo, na tabela, o esquema A possui um valor de categoria máximo de 3 e o esquema B possui um valor de categoria máximo de 7, mas ambos os esquemas codificam as mesmas três categorias.

O intervalo variável determina quais categorias serão omitidas a partir da análise. Quaisquer categorias com códigos fora do intervalo definido são omitidas da análise. Este é um método simples para a omissão de categorias, mas pode resultar em análises indesejadas. Uma categoria máxima definida incorretamente pode omitir categorias *válidas* a partir da análise. Por exemplo, para o esquema exemplo, definir o valor de categoria máximo para ser 3 indica que o *cargo* possui categorias codificadas de 1 a 3; a categoria *gerente* é tratada como ausente. Porque nenhuma categoria foi efetivamente codificada como 3, a

terceira categoria na análise não contém caso. Se o seu desejo era omitir todas as categorias de gerente, esta análise seria apropriada. No entanto, se os gerentes devem ser incluídos, a categoria máxima deve ser definida como 7, e os valores omissos devem ser codificados com valores acima de 7 ou abaixo de 1.

Para as variáveis tratadas como nominal ou ordinal, o intervalo das categorias não afeta os resultados. Para as variáveis nominais, apenas o rótulo e não o valor associado a esse rótulo é importante. Para as variáveis ordinais, a ordem das categorias é preservada nas quantificações; os valores de categoria em si não são importantes. Todos os esquemas de codificação resultantes na mesma ordenação de categoria terão resultados idênticos. Por exemplo, os primeiros três esquemas na tabela são funcionalmente equivalentes se *cargofor* analisado em um nível ordinal. A ordem das categorias é idêntica nesses esquemas. O esquema D, por outro lado, inverte a segunda e terceira categorias e produzirá resultados diferentes dos outros esquemas.

Embora muitos esquemas de codificação para uma variável sejam funcionalmente equivalentes, esquemas com diferenças pequenas entre os códigos são preferenciais, porque os códigos têm um impacto sobre a quantidade de saídas produzidas por um procedimento. Todas as categorias codificadas com valores entre 1 a o máximo definido pelo usuário são válidas. Se qualquer uma dessas categorias estão vazias, as quantificações correspondentes serão omissas do sistema ou 0, dependendo do procedimento. Embora nenhuma dessas designações afete as análises, a saída é produzida para essas categorias. Assim, para o esquema B, o *cargo* possui quatro categorias que recebem valores omissos do sistema. Para o esquema C, também há quatro categorias que recebem indicadores omissos do sistema. Em contraste, para o esquema A não há quantificações omissas do sistema. Usar números inteiros consecutivos como códigos para as variáveis tratadas como nominais ou ordinais resulta em muito menos saída sem afetar os resultados.

Os esquemas de codificação para as variáveis tratadas como numéricas são mais restritos que o caso ordinal. Para essas variáveis, as diferenças entre as categorias consecutivas são importantes. A tabela a seguir exibe três esquemas de codificação para *idade*.

Tabela 3. Esquemas de codificação alternativos para idade

Categoria	A	B	C
20	20	1	1
22	22	3	2
25	25	6	3
27	27	8	4

Qualquer recodificação de variáveis numéricas deve preservar as diferenças entre as categorias. Usar os valores originais é um método para assegurar a preservação das diferenças. No entanto, isso pode resultar em muitas categorias tendo os indicadores omissos do sistema. Por exemplo, o esquema A emprega os valores originais observados. Para todos os procedimentos de Categorias, exceto para Análise de Correspondência, o valor máximo da categoria é 27 e o valor mínimo da categoria é definido como 1. As primeiras 19 categorias são vazias e recebem os indicadores omissos do sistema. A saída pode rapidamente tornar-se bastante complicada se a categoria máxima for muito maior do que 1 e se houver muitas categorias vazias entre 1 e o máximo.

Para reduzir a quantidade de saída, uma recodificação pode ser feita. No entanto, no caso numérico, o recurso Recodificação Automática não deve ser usado. A codificação para números inteiros consecutivos resulta em diferenças de 1 entre todas as categorias consecutivas e, como resultado, todas as quantificações serão igualmente espaçadas. As características da métrica consideradas importantes ao tratar uma variável como numérica são destruídas pela recodificação para números inteiros consecutivos. Por exemplo, o esquema C na tabela corresponde ao registro automático de *idade*. A diferença entre as categorias 22 e 25 foi alterada de três para um, e as quantificações refletirão esta última diferença.

Um esquema de recodificação alternativa que preserva as diferenças entre as categorias é subtrair o menor valor da categoria de cada categoria e adicionar 1 para cada diferença. O esquema B resulta desta

transformação. O menor valor de categoria, 20, foi subtraído de cada categoria e 1 foi incluído em cada resultado. Os códigos transformados têm um mínimo de 1 e todas as diferenças são idênticas aos dados originais. O valor de categoria máximo é agora 8 e as quantificações zero antes da primeira quantificação não zero serão todas eliminadas. No entanto, as quantificações não zero correspondentes a cada categoria resultante do esquema B são idênticas às quantificações do esquema A.

Qual Procedimento É Melhor para o Seu Aplicativo?

As técnicas incorporadas em quatro destes procedimentos (Análise de correspondência, Análise de múltipla correspondência, Análise de componentes principais para dados categóricos e Análise de correlação canônica não linear) encaixam-se na área geral de análise de dados multivariados, conhecida como **redução de dimensão**. Ou seja, os relacionamentos entre as variáveis são representados em algumas dimensões—dois ou três— o mais rápido possível. Isto permite descrever estruturas ou padrões nos relacionamentos que seriam muito difíceis de entender em sua riqueza e complexidade inicial. Nos aplicativos de pesquisa de mercado, essas técnicas podem ser uma forma de **mapeamento perceptual**. Uma grande vantagem destes procedimentos é que eles acomodam dados com diferentes níveis de ajuste de escala ideal.

A regressão categórica descreve o relacionamento entre a variável de resposta categórica e uma combinação de variáveis preditoras categóricas. A influência de cada variável preditora na variável de resposta é descrita pela ponderação de regressão correspondente. Como em os outros procedimentos, os dados podem ser analisados com diferentes níveis de ajuste de escala ideal.

Escala multidimensional e Desdobramento multidimensional descrevem os relacionamentos entre objetos em um espaço de baixa dimensão, usando as proximidades entre os objetos.

A seguir, breves orientações para cada um dos procedimentos:

- Use a Regressão categórica para prever os valores de uma variável dependente categórica a partir de uma combinação de variáveis independentes categóricas.
- Use a Análise de componentes principais para dados categóricos para contabilizar os padrões de variação em um único conjunto de variáveis de níveis de ajuste de escala ideais mistos.
- Use a Análise de correlação canônica não linear para avaliar a extensão para a qual dois ou mais conjuntos de variáveis de níveis de ajuste de escala ideais mistos são correlacionados.
- Use a Análise de correspondência para analisar tabelas de contingência de duas vias ou dados que podem ser expressos como uma tabela de dois fatores, como preferência de marca ou dados de escolha sociométricos.
- Use a Análise de múltipla correspondência para analisar uma matriz categórica de dados multivariados quando estiver disposto a não fazer pressupostos mais fortes que todas as variáveis que são analisadas no nível nominal.
- Use o Ajuste de escala multidimensional para analisar dados de proximidade para localizar uma representação de quadrados mínimos de um conjunto de objetos único em um espaço de baixa dimensão.
- Use o Desdobramento multidimensional para analisar dados de proximidade para localizar uma representação de quadrados mínimos de dois conjuntos de objetos em um espaço de baixa dimensão.

Regressão categórica

O uso da Regressão Categórica é mais apropriada quando o objetivo de sua análise é prever uma variável dependente (resposta) a partir de um conjunto de variáveis independentes (preditor). Como acontece com todos os procedimentos de ajuste de escala ideais, valores da escala são atribuídos a cada uma das categorias de cada variável, de modo a que estes valores são ideais no que diz respeito à regressão. A solução de uma regressão categórica maximiza a correlação quadrada entre a resposta transformada e combinação ponderada de preditores transformados.

Relação com outros procedimentos de Categorias. A regressão categórica com ajuste de escala ideal é comparável à análise de correlação canônica de ajuste de escala ideal com dois conjuntos, um dos quais

que contém apenas a variável dependente. Na última técnica, a similaridade de conjuntos é derivada, comparando cada conjunto para uma variável desconhecida que se encontra em algum lugar entre todos os conjuntos. Na regressão categórica, a similaridade da resposta transformada e a combinação linear dos preditores transformados é avaliada diretamente.

Relação com técnicas padrão. Na regressão linear padrão, as variáveis categóricas podem ser recodificadas como variáveis indicadoras ou ser tratadas da mesma forma que as variáveis de nível de intervalo. Na primeira abordagem, o modelo contém um intercepto separado e uma inclinação para cada combinação dos níveis das variáveis categóricas. Isso resulta em um grande número paramétricas para interpretar. Na segunda abordagem, apenas um parâmetro é estimado para cada variável. No entanto, a natureza arbitrária das codificações de categoria torna as generalizações impossíveis.

Se alguma das variáveis não for contínua, as análises alternativas estarão disponíveis. Se a resposta for contínua e os preditores categóricos, a análise de variância será frequentemente empregada. Se a resposta for categórica e os preditores contínuos, a regressão logística ou a análise discriminante podem ser apropriadas. Se a resposta e os preditores forem ambos categóricos, os modelos log-linear serão frequentemente usados.

A regressão com ajuste de escala ideal oferece três níveis de ajuste de escala para cada variável. As combinações desses níveis podem contar para uma ampla gama de relações não lineares para os quais um único método "padrão" é mal adaptado. Consequentemente, o ajuste de escala ideal oferece maior flexibilidade do que a abordagem padrão com complexidade mínima incluída.

Além disso, as transformações não lineares dos preditores geralmente reduzem as dependências entre os preditores. Se você comparar os autovalores da matriz de correlação para os preditores com os autovalores da matriz de correlação para os preditores com ajuste de escala ideal, o último conjunto será geralmente menos variável do que o primeiro. Em outras palavras, na regressão categórica, o ajuste de escala ideal torna os maiores autovalores da matriz de correlação do preditor menores e os menores autovalores maiores.

Análise de componentes principais para dados categóricos

O uso da Análise de componentes principais para dados categóricos é mais apropriada quando se deseja contabilizar os padrões de variação em um único conjunto de variáveis de níveis de ajuste de escala ideais mistos. Esta técnica tenta reduzir a dimensionalidade de um conjunto de variáveis enquanto considera o máximo de variação possível. Os valores de escala são designados a cada categoria de cada variável, de modo que estes valores são ideais com respeito à solução dos principais componentes. Os objetos na análise recebem pontuações de componentes baseados nos dados quantificados. Os gráficos das pontuações dos componentes revelam padrões entre os objetos na análise e podem revelar objetos incomuns nos dados. A solução de uma análise de componentes principais para dados categóricos maximiza as correlações das pontuações de objetos com cada uma das variáveis quantificadas para o número de componentes (dimensões) especificados.

Uma aplicação importante dos componentes principais categóricos é examinar os dados de preferência, nos quais os respondentes classificam ou avaliam uma série de itens com respeito a preferência. Na configuração de dados normal do IBM SPSS Statistics, as linhas são individuais, as colunas são medições para os itens e as pontuações entre as linhas são as pontuações de preferência (em uma escala de 0 a 10, por exemplo), tornando os dados condicionais por linha. Para dados de preferência, você pode desejar tratar os indivíduos como variáveis. Usando o procedimento Transpor, é possível transpor os dados. Os avaliadores transformam as variáveis, e todas as variáveis são declaradas ordinais. Não há objeção para usar mais variáveis que objetos em CATPCA.

Relação com outros procedimentos de Categorias. Se todas as variáveis forem declaradas múltiplas nominais, a análise de componentes principais para dados categóricos produzirá uma análise equivalente para uma análise de múltipla correspondência ser executada nas mesmas variáveis. Assim, a análise de componentes principais para dados categóricos pode ser vista como um tipo de análise de múltipla correspondência na qual algumas das variáveis são declaradas ordinais ou numéricas.

Relação com técnicas padrão. Se todas as variáveis são dimensionadas no nível numérico, a análise de componentes principais para dados categóricos é equivalente a análise de componentes principais padrão.

Em termos mais gerais, a análise de componentes principais para dados categóricos é uma alternativa ao cálculo das correlações entre escalas não numéricas e analisá-las usando uma abordagem de componentes principais padrão ou análise de fator. O uso sem malícia do coeficiente de correlação de Pearson normal como medida de associação para dados ordinais pode levar ao viés não trivial na estimação das correlações.

Análise de correlação canônica não linear

A Análise de correlação canônica não linear é um procedimento bem geral com muitos aplicativos diferentes. O objetivo da análise de correlação canônica não linear é analisar os relacionamentos entre dois ou mais conjuntos de variáveis, em vez de entre as próprias variáveis, como na análise de componentes principais. Por exemplo, é possível ter dois conjuntos de variáveis, onde um conjunto de variáveis pode ser de itens de plano de fundo demográficos sobre um conjunto de respondentes e um segundo conjunto pode ser de respostas a um conjunto de itens de atitude. Os níveis de ajuste de escala na análise podem ser qualquer combinação de nominal, ordinal e numérico. A análise correlacional canônica de ajuste de escala ideal determina a semelhança entre os conjuntos, comparando simultaneamente as variáveis canônicas de cada conjunto para um conjunto de compromisso das pontuações atribuídas aos objetos.

Relação com outros procedimentos de Categorias. Se existirem dois ou mais conjuntos de variáveis com apenas uma variável por conjunto, a análise de correlação canônica de ajuste de escala ideal será equivalente a análise de componentes principais de ajuste de escala ideal. Se todas as variáveis em uma análise de uma variável por conjunto forem multinominais, a análise de correlação canônica de ajuste de escala ideal será equivalente a análise múltipla correspondência. Se houver dois conjuntos de variáveis, um deles contendo uma única variável, a análise de correlação canônica de ajuste de escala ideal é equivalente a regressão categórica com ajuste de escala ideal.

Relação com técnicas padrão. A análise de correlação canônica padrão é uma técnica estatística que encontra uma combinação linear de um conjunto de variáveis e uma combinação linear de um segundo conjunto de variáveis que são maximamente correlacionados. Dado este conjunto de combinações lineares, a análise de correlação canônica pode encontrar conjuntos independentes de combinações lineares subsequentes, que se referem a variáveis canônicas, até um número máximo igual ao número de variáveis no conjunto menor.

Se houver dois conjuntos de variáveis na análise e todas as variáveis forem definidas como numéricas, a análise de correlação canônica de ajuste de escala ideal será equivalente a uma análise de correlação canônica padrão. Embora IBM SPSS Statistics não tenha um procedimento de análise de correlação canônica, muitas estatísticas relevantes podem ser obtidas da análise de variação multivariada.

A análise de correlação canônica de ajuste de escala ideal possui várias outras aplicações. Se você tem dois conjuntos de variáveis e um dos conjuntos contém uma variável nominal declarada como única nominal, os resultados da análise de correlação canônica de ajuste de escala ideal poderão ser interpretados de forma semelhante a análise de regressão. Se você considerar que a variável seja multinominal, a análise de ajuste de escala ideal será uma alternativa a análise discriminante. O agrupamento das variáveis em mais de dois conjuntos fornece uma variedade de maneiras de analisar seus dados.

Análise de correspondências

O objetivo da análise de correspondência é fazer biplots para tabelas de correspondência. Em uma tabela de correspondência, as variáveis de linha e coluna são assumidas para representar as categorias não ordenadas; portanto, o nível de ajuste de escala ideal nominal é sempre usado. Ambas as variáveis são inspecionadas para suas informações nominais apenas. Isto é, a única consideração é o fato de alguns objetos estarem na mesma categoria, enquanto outros não estão. Nada é assumido sobre a distância ou a ordem entre as categorias da mesma variável.

Um uso específico da análise de correspondência é a análise de tabelas de contingência de dois fatores. Se uma tabela tem linhas r ativas e colunas c ativas, o número de dimensões na solução de análise de correspondência é o mínimo de r menos 1 ou c menos 1, seja qual for o menor. Em outras palavras, é possível representar perfeitamente as categorias de linha ou as categorias de colunas de uma tabela de contingência em um espaço de dimensões. Em termos práticos, no entanto, você gostaria de representar as categorias de linha e colunas de uma tabela de dois fatores em um espaço de baixa dimensão, ou seja, duas dimensões, pela razão de que os gráficos bidimensionais são mais facilmente compreensíveis do que representações espaciais multidimensionais.

Quando menos do que o número máximo de possíveis dimensões é usado, as estatísticas produzidas na análise descrevem quão bem as categorias de linhas e colunas são representadas na representação de baixa dimensão. Desde que a qualidade da representação da solução bidimensional seja boa, será possível examinar gráficos dos pontos de linha e os pontos de coluna para saber quais categorias da variável de linha são semelhantes, quais categorias de variável de coluna são semelhantes e quais categorias de linha e coluna são semelhantes umas às outras.

Relação com outros procedimentos de Categorias. A análise de correspondência é limitada a tabelas de dois fatores. Se houver mais de duas variáveis de interesse, será possível combinar variáveis para criar variáveis de interação. Por exemplo, para as variáveis *region*, *job* e *age*, é possível combinar *region* e *job* para criar uma nova variável *rejob* com as 12 categorias mostradas na tabela a seguir. Esta nova variável forma uma tabela de dois fatores com *age* (12 linhas, 4 colunas), que pode ser analisada na análise de correspondência.

Tabela 4. Combinações de região e cargo

Código de categoria	Definição de categoria	Código de categoria	Definição de categoria
1	Norte, estagiário	7	Leste, estagiário
2	Norte, representante de vendas	8	Leste, representante de vendas
3	Norte, gerente	9	Leste, gerente
4	Sul, estagiário	10	Oeste, estagiário
5	Sul, representante de vendas	11	Oeste, representante de vendas
6	Sul, gerente	12	Oeste, gerente

Uma desvantagem desta abordagem é que todos os pares de variáveis podem ser combinados. É possível combinar *cargo* e *idade*, produzindo outra variável de 12 categorias. Ou é possível combinar *região* e *idade*, que resulta em uma nova variável de 16 categorias. Cada uma dessas variáveis de interação forma uma tabela de dois fatores com a variável restante. As análises de correspondência dessas três tabelas não produzirão resultados, mas cada uma é uma abordagem válida. Além disso, se houver quatro ou mais variáveis, tabelas de dois fatores comparando uma variável de interação com outra variável de interação poderão ser construídas. O número de possíveis tabelas para analisar pode ficar muito grande, mesmo para algumas variáveis. É possível selecionar uma dessas tabelas para analisar, ou é possível analisar todas elas. Alternativamente, o procedimento Análise de múltipla correspondência pode ser usado para examinar todas as variáveis simultaneamente, sem a necessidade de construir variáveis de interação.

Relação com técnicas padrão. O procedimento Tabulações cruzadas pode também ser usado para analisar tabelas de contingências, com a independência como um foco comum na análise. No entanto, mesmo em tabelas pequenas, detectar a causa de partidas a partir de independência pode ser difícil. O utilitário de análise de correspondência encontra-se em exibição de tais padrões para tabelas de dois fatores de qualquer tamanho. Se existir uma associação entre as variáveis de linha e coluna, isto é, se o valor qui-quadrado é significativo, a análise de correspondência poderá ajudar a revelar a natureza do relacionamento.

Análise de correspondências múltipla

A Análise de múltipla correspondência tenta produzir uma solução na qual os objetos dentro da mesma categoria são plotados próximos e objetos de diferentes categorias são plotados distantes. Cada objeto fica tão perto quanto possível dos pontos de categoria das categorias que são aplicados ao objeto. Dessa forma, as categorias dividem os objetos em subgrupos homogêneos. As variáveis são consideradas homogêneas quando classificam os objetos nas mesmas categorias para os mesmos subgrupos.

Para uma solução unidimensional, a análise de múltipla correspondência atribui valores de escala ideais (quantificações de categoria) para cada categoria de cada variável, de modo que, geralmente, em média, as categorias têm máxima difusão. Para uma solução bidimensional, a análise de múltipla correspondência encontra um segundo conjunto de quantificações das categorias de cada variável não relacionada ao primeiro conjunto, tentando maximizar novamente a difusão, e assim por diante. Porque as categorias de uma variável recebem tantas pontuações quanto existem dimensões, as variáveis na análise são assumidas como nominais múltiplas no nível do ajuste de escala ideal.

A análise de múltipla correspondência também atribui pontuações aos objetos na análise, de tal modo que as quantificações de categoria são as médias, ou centroides, dos escores de objeto dos objetos nessa categoria.

Relação com outros procedimentos de Categorias. A análise de múltipla correspondência também é conhecida como análise de homogeneidade ou ajuste de escala dual. Ele dá resultados comparáveis, mas não idênticos, para análise de correspondência quando há apenas duas variáveis. A análise de correspondência produz uma saída única, resumindo a forma e qualidade da representação da solução, que inclui informações sobre a estabilidade. Assim, a análise de correspondência é geralmente preferível à análise de múltipla correspondência no caso de duas variáveis. Outra diferença entre os dois procedimentos é que a entrada para a análise de múltipla correspondência é uma matriz de dados, onde as linhas são objetos e as colunas são variáveis, enquanto que a entrada para análise de correspondência pode ser a mesma matriz de dados, uma matriz de proximidade geral, ou uma tabela de contingência conjunta, que é uma matriz agregada na qual ambas as linhas e colunas representam categorias de variáveis. A análise de múltipla correspondência também pode ser pensada como uma análise de componentes principais dos dados em escala no nível nominal múltiplo.

Relação com técnicas padrão. A análise de múltipla correspondência pode ser pensada como a análise de uma tabela de contingência multiponto. As tabelas de contingência multipontos também podem ser analisadas com o procedimento Tabulações cruzadas, mas este procedimento fornece estatísticas básicas separadas para cada categoria de cada variável de controle. Com a análise de múltipla correspondência, muitas vezes é possível resumir o relacionamento entre todas as variáveis com um único gráfico bidimensional. Um uso avançado de análise de múltipla correspondência é substituir os valores de categoria originais com os valores de escala ideal a partir da primeira dimensão e executar uma análise multivariada secundária. Como a análise de múltipla correspondência substitui os rótulos de categoria por valores numéricos de escala, muitos procedimentos diferentes que requerem dados numéricos podem ser aplicados após a análise de múltipla correspondência. Por exemplo, o procedimento Análise Fatorial produz um primeiro componente principal que é equivalente à primeira dimensão de análise de múltipla correspondência. Os escores dos componentes na primeira dimensão são iguais aos escores do objeto, e os carregamentos de componentes quadrados são iguais para as medidas de discriminação. A segunda dimensão de análise de múltipla correspondência, no entanto, não é igual à segunda dimensão da análise fatorial.

Escala multidimensional

O uso do Ajuste de escala multidimensional é mais apropriado quando o objetivo de sua análise é localizar a estrutura em um conjunto de medidas de distância entre um único conjunto de objetos ou casos. Isto é executado ao designar observações para especificar localizações em um espaço conceitual de baixa dimensão, de modo que as distâncias entre pontos no espaço correspondam à determinadas (dis)similaridades tão próximas quanto possível. O resultado é uma representação de quadrados mínimos dos objetos nesse espaço de baixa dimensão, que, em muitos casos, vai ajudá-lo a entender melhor seus dados.

Relação com outros procedimentos de Categorias. Quando você tem dados multivariados a partir dos quais você cria distâncias e que você, em seguida, analisa com ajuste de escala multidimensional, os resultados serão semelhantes para analisar os dados utilizando a análise de componentes principais para dados categóricos com normalização principal de objeto. Este tipo de análise de componentes principais também é conhecida como análise de coordenadas principais.

Relação com técnicas padrão. O procedimento Ajuste de escala multidimensional de categorias (PROXSCAL) oferece diversas melhorias sobre o procedimento de ajuste de escala disponível no Statistics Base Edition (ALSCAL). O PROXSCAL oferece um algoritmo acelerado para determinados modelos e permite colocar restrições no espaço comum. Além disso, o PROXSCAL tenta minimizar o estresse bruto normalizado, em vez de S-stress (também referido como **strain**). O estresse bruto normalizado é geralmente preferencial ele é uma medida baseada nas distâncias, enquanto o S-stress é baseado nas distâncias quadradas.

Desdobramento multidimensional

O uso do Desdobramento multidimensional é mais apropriado quando o objetivo de sua análise é localizar a estrutura em um conjunto de medidas de distância entre dois conjuntos de objetos (referenciados como objetos de linha e coluna). Isto é executado ao designar observações para especificar localizações em um espaço conceitual de baixa dimensão, de modo que as distâncias entre pontos no espaço correspondam à determinadas (dis)similaridades tão próximas quanto possível. O resultado é uma representação de quadrados mínimos dos objetos de linha e coluna nesse espaço de baixa dimensão, que, em muitos casos, vai ajudá-lo a entender melhor seus dados.

Relação com outros procedimentos de Categorias. Se seus dados consistirem de distâncias entre um conjunto único de objetos (um quadrado, matriz simétrica), use o ajuste de escala multidimensional.

Relação com técnicas padrão. O Desdobramento multidimensional de categorias (PREFSCAL) oferece várias melhorias sobre a funcionalidade de desdobramento disponível no Statistics Base Edition (por meio de ALSCAL). PREFSCAL permite inserir restrições no espaço comum; além disso, PREFSCAL tenta minimizar uma medida de estresse penalizado que ajuda a evitar soluções degeneradas (as quais algoritmos mais velhos são propensos).

Proporção de aspecto nos gráficos de ajuste de escala ideal

A proporção de aspecto nos gráficos de ajuste de escala ideal são isotrópicos. Em um gráfico bidimensional, a distância que representa uma unidade na dimensão 1 é igual a distância que representa uma unidade na dimensão 2. Se você alterar o intervalo de uma dimensão em um gráfico bidimensional, o sistema muda o tamanho da outra dimensão para manter as distâncias físicas iguais. A proporção de aspecto isotrópica não pode ser substituída para os procedimentos de ajuste de escala ideal.

Regressão categórica (CATREG)

A **Regressão categórica** quantifica dados categóricos designando valores numéricos para as categorias, o que resulta em uma equação de regressão linear ideal para as variáveis transformadas. A Regressão categórica também é conhecida pelo acrônimo CATREG, para *regressão categórica*.

A análise de regressão linear padrão envolve minimizar a soma das diferenças quadráticas entre uma variável de resposta (dependente) e uma combinação ponderada de variáveis preditoras (independentes). As variáveis são normalmente quantitativas, com dados categóricos (nominais) registrados para variáveis binárias ou de contraste. Como resultado, as variáveis categóricas servem para separar grupos de casos e a técnica estima conjuntos paramétricos separados para cada grupo. Os coeficientes estimados refletem como as mudanças feitas nos preditores afetam a resposta. A predição da resposta é possível para qualquer combinação de valores do preditor.

Uma abordagem alternativa envolve retornar a resposta nos próprios valores do preditor categórico. Consequentemente, um coeficiente é estimado para cada variável. No entanto, para variáveis categóricas,

os valores de categoria são arbitrários. Codificar as categorias de diferentes maneiras produz coeficientes diferentes, fazendo as comparações entre análises das mesmas variáveis ficarem difíceis.

CATREG amplia a abordagem padrão, ajustando simultaneamente as escalas das variáveis nominais, ordinais e numéricas. O procedimento quantifica variáveis categóricas, de modo que as quantificações refletem características das categorias originais. O procedimento trata variáveis categóricas quantificadas da mesma forma que as variáveis numéricas. Usar as transformações não lineares permite que as variáveis sejam analisadas em uma variedade de níveis, a fim de localizar o modelo de melhor ajuste.

Exemplo. A regressão categórica pode ser usada para descrever como a satisfação do cargo depende da categoria do cargo, região geográfica e quantia de viagens. É possível descobrir que os altos níveis de satisfação correspondem a gerentes e baixas de viagens. A equação de regressão resultante pode ser usada para prever a satisfação do cargo para qualquer combinação das três variáveis independentes.

Estatísticas e gráficos. Frequências, coeficientes de regressão, tabela ANOVA, histórico de iteração, quantificações de categoria, correlações entre preditores não transformados, correlações entre preditores transformados, gráficos de resíduos e gráficos de transformação.

Considerações sobre dados de regressão categórica

Dados. O CATREG opera em variáveis de indicadores de categoria. Os indicadores de categoria devem ser números inteiros positivos. É possível usar a caixa de diálogo Discretização para converter variáveis de valor fracionário e variáveis de sequência em números inteiros positivos.

Suposições. Somente uma variável de resposta é permitida, mas o número máximo de variáveis preditoras é 200. Os dados devem conter pelo menos três casos válidos, e o número de casos válidos deve exceder o número de variáveis preditoras mais um.

Procedimentos relacionados. CATREG é equivalente à análise de correlação canônica de dados categóricos com ajuste de escala ideal (OVERALS) com dois conjuntos, um dos quais contendo apenas uma variável. O ajuste de escala de todas as variáveis no nível numérico corresponde a análise de regressão múltipla padrão.

Para obter uma regressão categórica

1. Nos menus, escolha:
Analisar > Regressão > Ajuste de escala ideal (CATREG)...
2. Selecione a variável dependente e variável(is) independente(s).
3. Clique em **OK**.

Opcionalmente, altere o nível de ajuste de escala para cada variável.

Definir escala na regressão categórica

É possível configurar o nível de ajuste de escala ideal para as variáveis dependentes e independentes. Por padrão, elas são escalas como splines monotônicas de segundo grau (ordinal) com dois nós internos. Além disso, é possível configurar a ponderação para variável de análise.

Nível de ajuste de escala ideal. Também é possível selecionar o nível de ajuste de escala para quantificar cada variável.

- **Ordinal spline.** A ordem das categorias da variável observada é preservada na variável idealmente escalada. Os pontos de categoria estarão em uma linha reta (vetor) até a origem. A transformação resultante é um polinomial monotônico suave em trechos do grau escolhido. As partes são especificadas pelo número especificado pelo usuário e pela colocação determinada pelo procedimento dos nós internos.

- **Nominal spline.** A única informação na variável observada que é preservada na variável idealmente escalada é o agrupamento de objetos em categorias. A ordem das categorias da variável observada não é preservada. Os pontos de categoria estarão em uma linha reta (vetor) até a origem. A transformação resultante é um polinomial não monotônico suave em trechos do grau escolhido. As partes são especificadas pelo número especificado pelo usuário e pela colocação determinada pelo procedimento dos nós internos.
- **Ordinal.** A ordem das categorias da variável observada é preservada na variável escalada de modo ideal. Os pontos de categoria estarão em uma linha reta (vetor) até a origem. A transformação resultante se ajusta melhor que a transformação ordinal de spline, mas é menos suave.
- **Nominal.** A única informação na variável observada que é preservada na variável idealmente escalada é o agrupamento de objetos em categorias. A ordem das categorias da variável observada não é preservada. Os pontos de categoria estarão em uma linha reta (vetor) até a origem. A transformação resultante se ajusta melhor que a transformação nominal de spline, mas é menos suave.
- **Númerico.** As categorias são tratadas conforme ordenadas e igualmente espaçadas (nível de intervalo). A ordem das categorias e as distâncias iguais entre os números de categoria da variável observada são preservadas na variável idealmente escalada. Os pontos de categoria estarão em uma linha reta (vetor) até a origem. Quando todas as variáveis estão no mesmo nível numérico, a análise é análoga à análise de componentes principais padrão.

Discretização de regressão categórica

A caixa de diálogo Discretização permite selecionar um método de gravar suas variáveis. As variáveis de valor fracionário são agrupadas em sete categorias (ou no número de valores distintos da variável, se este número for menor que sete) com uma distribuição aproximadamente normal, a menos que seja especificado de outra forma. As variáveis de sequência são sempre convertidas em números inteiros positivos, designando indicadores de categoria de acordo com a ordem alfanumérica ascendente. A discretização para variáveis de sequência é aplicada a esses números inteiros. Outras variáveis são deixadas sozinhas por padrão. As variáveis discretizadas são, então, usadas na análise.

Método. Escolha entre agrupamento, ranqueamento e multiplicação.

- **Agrupamento.** Recodifique para um número especificado de categorias ou recodifique por intervalo.
- **Ranqueamento.** A variável é discretizada pelo pelos postos dos casos.
- **Multiplicação.** Os valores atuais da variável são padronizados, multiplicados por 10, arredondados e têm uma constante incluída, de modo que o menor valor discretizado é 1.

Agrupamento. Os pontos a seguir estão disponíveis ao discretizar variáveis por agrupamento:

- **Número de categorias.** Especifique o número de categorias e se os valores da variável devem seguir uma distribuição aproximadamente normal ou uniforme entre essas categorias.
- **Intervalos iguais.** As variáveis são recodificadas para categorias definidas por esses intervalos dimensionados igualmente. Deve-se especificar o comprimento dos intervalos.

Valores omissos da regressão categórica

A caixa de diálogo Valores omissos permite escolher a estratégia para tratamento de valores omissos em variáveis de análise e variáveis complementares.

Estratégia. Escolha excluir objetos com valores omissos (exclusão de listwise) ou inserir valores omissos (tratamento ativo).

- **Exclua os objetos com valores omissos nesta variável.** Os objetos com valores omissos na variável selecionada são excluídos da análise. Esta estratégia não está disponível para variáveis complementares.
- **Imputar valores omissos.** Os objetos com valores omissos na variável selecionada têm estes valores imputados. É possível escolher o método de imputação. Selecione **Modo** para substituir valores omissos com a categoria mais frequente. Quando houver vários nós, aquele com o indicador de

categoria menor será usado. Selecione **Categoria extra** para substituir valores omissos pela mesma quantificação de uma categoria extra. Isto implica que objetos com um valor omissos nesta variável são considerados para pertencer à mesma categoria (extra).

Opções de regressão categórica

A caixa de diálogo Opções permite selecionar o estilo de configuração inicial, especificar critérios de iteração e convergência, selecionar objetos complementares e configurar a rotulagem de gráficos.

Objetos complementares. Isto permite especificar os objetos que você deseja tratar como complementar. Basta digitar o número de um objeto complementar (ou especificar um intervalo de casos) e clicar em **Incluir**. Não é possível ponderar objetos complementares (ponderações especificadas são ignoradas).

Configuração inicial. Se nenhuma variável for tratada, selecione a configuração **Numérica**. Se pelo menos uma variável for tratada como nominal, selecione a configuração **Aleatória**.

Como alternativa, se pelo menos uma variável tiver um nível de ajuste de escala ordinal ou ordinal spline, o algoritmo de ajuste de modelo comum poderá resultar em uma solução subideal. Escolher **Múltiplos inícios sistemáticos** com todos os possíveis padrões de sinais a serem testados sempre encontrará a solução ideal, mas o tempo de processamento necessário aumenta rapidamente conforme o número de variáveis ordinais e ordinais spline no conjunto de dados aumenta. É possível reduzir o número de padrões de teste especificando uma porcentagem de perda de limite de variância, em que o quanto maior o limiar, mais padrões de sinal serão excluídos. Com esta opção, obter a solução ideal não é garantido, mas a chance de obter uma solução subideal é diminuída. Além disso, se a solução ideal não for encontrada, a chance de que a solução subideal seja muito diferente da solução ideal é diminuída. Quando vários inícios sistemáticos são solicitados, os sinais dos coeficientes de regressão para cada início são gravados em um arquivo de dados ou conjunto de dados externo do IBM SPSS Statistics na sessão atual. Consulte o tópico "Salvamento de regressão categórica" na página 15 para obter mais informações

Os resultados de uma execução anterior com vários inícios sistemáticos permite **Usar sinais fixos para os coeficientes de regressão**. Os sinais (indicado por 1 e -1) precisam estar em uma linha do conjunto de dados ou arquivo especificado. O número de partida de valores de número inteiro é o número do caso da linha neste arquivo que contém os sinais a serem usados.

Crítérios. É possível especificar o número máximo de interações que a regressão pode percorrer em seus cálculos. Também é possível selecionar um valor de critério de convergência. A regressão interrompe a iteração se a diferença no ajuste total entre as últimas duas iterações for menor que o valor de convergência ou se o número máximo de iterações for atingido.

Gráficos por rótulo. Permite especificar se as variáveis e os rótulos de valor ou nomes e valores de variáveis serão usados nos gráficos. Também é possível especificar um comprimento máximo para rótulos.

Rgularização de regressão categórica

Método. Os métodos de regularização podem melhorar o erro preditivo do modelo, reduzindo a variabilidade nas estimativas de coeficiente de regressão, reduzindo as estimativas em direção a 0. O Laço e a Rede Elástica encolherao algumas estimativas de coeficientes para exatamente 0, fornecendo, assim, uma forma de seleção de variáveis. Quando um método de regularização é solicitado, o modelo regularizado e os coeficientes para cada valor de coeficiente de penalidade são gravados em um arquivo de dados externo do IBM SPSS Statistics ou em um conjunto de dados na sessão atual. Consulte o tópico "Salvamento de regressão categórica" na página 15 para obter mais informações

- **Enrugar regressão.** Enrugar regressão reduz os coeficientes introduzindo um termo de penalidade igual à soma de coeficientes quadrados vezes um **coeficiente de penalidade**. Este coeficiente pode variar de 0 (nenhuma penalidade) a 1; o procedimento procurará o "melhor" valor da penalidade se você especificar um intervalo e incremento.

- **Laço.** O termo de penalidade do Laço é baseado na soma dos coeficientes absolutos, e a especificação de um coeficiente de penalidade é semelhante ao da regressão de Ridge; no entanto, o Laço é mais intensivo computacionalmente.
- **Rede elástica.** A Rede Elástica combina simplesmente as penalidades da regressão Laço e Ridge e procurará na grade de valores especificados para encontrar os "melhores" coeficientes de penalidade de regressão de Lasso e de Ridge. Para um determinado par de penalidades de regressão de Laço e Ridge, a Rede Elástica não é muito mais computacionalmente caro que o Laço.

Exibir gráficos de regularização. Esses são os gráficos de coeficiente de regressão versus a penalidade de regularização. Ao procurar um intervalo de valores para o "melhor" coeficiente de regressão, ele fornece uma visualização de como os coeficientes de regressão mudam ao longo desse intervalo.

Gráficos de rede elástica. Para o método Rede Elástica, gráficos de regularização separados são produzidos por valores da penalidade de regressão de Ridge. **Todos os possíveis gráficos** usam cada valor no intervalo determinado pelos valores de penalidade mínimo e máximo da regressão de Ridge especificados. **Para algumas penalidades de Ridge** permite especificar um subconjunto dos valores no intervalo determinado pelo mínimo e máximo. Basta digitar o número de um valor de penalidade (ou especificar um intervalo de valores) e clicar em **Incluir**.

Saída de regressão categórica

A caixa de diálogo Saída permite selecionar as estatísticas a serem exibidas na saída.

Tabelas. Produz tabelas para:

- **R múltiplo.** Inclui R^2 , R^2 ajustado e R^2 justado levando o ajuste de escala ideal em consideração.
- **ANOVA.** Esta opção inclui regressão e somas de quadrados dos resíduos, quadrados médio e F . Duas novas tabelas ANOVA são exibidas: uma com graus de liberdade para a regressão igual ao número de variáveis preditoras e outra com graus de liberdade para a regressão que leva o ajuste de escala ideal em consideração.
- **Coefficientes.** Esta opção fornece três tabelas: uma tabela de coeficientes, que inclui betas, erros padrão dos betas, valores de t e significância; uma tabela de Ajuste de escala ideal de coeficientes, com o erro padrão dos betas levando em consideração os graus de liberdade do ajuste de escala ideal; e uma tabela com a correlação de ordem zero, parte e parcial, medida de importância relativa de Pratt e a tolerância antes e depois da transformação.
- **Histórico de iteração.** Para cada iteração, incluindo os valores de início para o algoritmo, o R múltiplo e o erro de regressão são mostrados. O aumento no R múltiplo é listado a partir da primeira iteração.
- **Correlações de variáveis originais.** Uma matriz mostrando a correlação entre as variáveis não transformadas é exibida.
- **Correlações de variáveis transformadas.** Uma matriz mostrando a correlação entre as variáveis transformadas é exibida.
- **Modelos e coeficientes regularizados.** Exibe valores de penalidade, quadrado-R e os coeficientes de regressão para cada modelo regularizado. Se um método de reamostragem for especificado ou se objetos complementares (casos de teste) forem especificados, ele também exibirá o erro de predição ou teste MSE.

Reamostragem. Os métodos de reamostragem dão uma estimativa do erro de predição do modelo.

- **Validação cruzada.** A tabulação cruzada divide a amostra em um número de subamostras, ou dobras. Os modelos de regressão categórica são, então, gerados, excluindo os dados de cada subamostra por vez. O primeiro modelo baseia-se em todos os casos, exceto aqueles na primeira dobra de amostra, o segundo modelo baseia-se em todos os casos, exceto aqueles na segunda dobra de amostra, e assim por diante. Para cada modelo, o erro de predição é estimado ao aplicar o modelo à subamostra excluída ao gerá-lo.
- **.632 Bootstrap.** Com o bootstrap, as observações são colhidas aleatoriamente a partir dos dados com substituição, repetindo este processo várias vezes para obter várias amostras bootstrap. Um modelo é

ajustado para cada amostra bootstrap. O erro de predição para cada modelo é estimado ao aplicar o modelo ajustado aos casos que não estão na amostra bootstrap.

Quantificações de categoria. As tabelas que mostram os valores transformados das variáveis selecionadas são exibidas.

Estatísticas descritivas. As tabelas que mostram as frequências, valores omissos e modos das variáveis selecionadas são exibidas.

Salvamento de regressão categórica

A caixa de diálogo Salvar permite salvar valores preditos, resíduos e valores transformados no conjunto de dados ativo e/ou salvar dados discretizados, valores transformados, modelos e coeficientes regularizados e sinais de coeficientes de regressão em um arquivo de dados ou conjunto de dados externo do IBM SPSS Statistics na sessão atual.

- Os conjuntos de dados estão disponíveis durante a sessão atual, mas não estão disponíveis nas sessões subsequentes, a menos que sejam salvos explicitamente como arquivos de dados. Os nomes dos conjuntos de dados devem aderir às regras de nomenclatura de variáveis.
- Os nomes dos arquivos ou os nomes do conjunto de dados devem ser diferentes para cada tipo de dado salvo.

Modelos e coeficientes regularizados são salvos sempre que um método de regularização é selecionado no diálogo Regularização. Por padrão, o procedimento cria um novo conjunto de dados com um nome exclusivo, mas é possível, naturalmente, especificar um nome de sua própria escolha ou gravar em um arquivo externo.

Os sinais de coeficientes de regressão são salvos sempre que vários inícios sistemáticos são usados como configuração inicial no diálogo Opções. Por padrão, o procedimento cria um novo conjunto de dados com um nome exclusivo, mas é possível, naturalmente, especificar um nome de sua própria escolha ou gravar em um arquivo externo.

Gráficos de transformação de regressão categórica

A caixa de diálogo Gráficos permite especificar as variáveis que produzirão transformação e gráficos de resíduos.

Gráficos de transformação. Para cada uma dessas variáveis, as quantificações de categoria são plotadas com relação aos valores de categoria original. Categorias vazias aparecem no eixo horizontal, mas não afetam os cálculos. Essas categorias são identificadas por quebras na linha que conecta as quantificações.

Gráficos de resíduos. Para cada uma dessas variáveis, os resíduos (calculados para a variável dependente predita a partir de todas as variáveis preditas, exceto a variável predita em questão) são plotados com relação aos indicadores de categoria e as quantificações de categoria ideal com beta em relação aos indicadores de categoria.

Recursos adicionais do comando CATREG

É possível customizar sua regressão categórica se você colar suas seleções em uma janela de sintaxe e editar a sintaxe de comando CATREG resultante. O idioma da sintaxe de comando também permite:

- Especifique nomes de raiz para as variáveis transformadas ao salvá-las para o conjunto de dados ativo (com o subcomando SAVE).

Consulte a *Referência da sintaxe de comando* para obter informações de sintaxe completa.

Análise de componentes principais para dados categóricos (CATPCA)

Este procedimento quantifica simultaneamente as variáveis categóricas enquanto reduz a dimensionalidade dos dados. A análise de componentes principais para dados categóricos também é conhecida pelo acrônimo CATPCA, para análise de componentes principais para dados *categóricos*.

O objetivo da análise de componentes principais é reduzir um conjunto original de variáveis em um conjunto menor de componentes não correlacionados que representa a maior parte das informações encontradas nas variáveis originais. A técnica é mais útil quando um grande número de variáveis proíbe interpretação eficaz das relações entre objetos (assuntos e unidades). Ao reduzir a dimensionalidade, você interpreta alguns componentes em vez de um grande número de variáveis.

A análise de componentes principais padrão assume relações lineares entre as variáveis numéricas. Por outro lado, a abordagem de ajuste de escala ideal permite que as variáveis sejam escaladas em diferentes níveis. As variáveis categóricas são quantificadas idealmente na dimensionalidade especificada. Como resultado, os relacionamentos não lineares entre as variáveis podem ser modelados.

Exemplo. A análise de componentes principais para dados categóricos poderia ser usada para apresentar graficamente a relação entre a categoria de cargo, divisão de cargo, região, quantidade de viagens (alta, média e baixa) e satisfação no trabalho. Você pode achar que duas dimensões são responsáveis por uma grande quantidade de variância. A primeira dimensão pode separar as categorias de cargos da região, ao passo que a segunda dimensão pode separar a divisão de cargos da quantidade de viagens. Também é possível achar que a alta satisfação no trabalho está relacionada a uma quantidade média de viagens.

Estatísticas e gráficos. Frequências, valores omissos, nível de ajuste de escala ideal, modo, variância contabilizada por coordenadas do centroide, coordenadas de vetor, total por variável e por dimensão, carregamentos de componente para variáveis quantificadas por vetores, quantificações e coordenadas de categoria, histórico de iteração, correlações das variáveis transformadas e autovalores da matriz de correlação, correlações das variáveis originais e autovalores da matriz de correlação, escores dos objetos, gráficos de categoria, gráficos de categorias conjuntas, gráficos de transformação, gráficos residuais, gráficos de centroide projetado, gráficos de objetos, biplots, triplots e gráficos de carregamentos de componentes.

Considerações de dados de análise de componentes principais para dados categóricos

Dados. Os valores de variável de sequência de caracteres são sempre convertidos em números inteiros positivos pela ordem alfanumérica ascendente. Valores omissos definidos pelo usuário, valores omissos do sistema e valores menores que 1 são considerados omissos; é possível recodificar ou incluir uma constante para variáveis com valores inferiores a 1 para torná-las não omissas.

Suposições. Os dados devem conter pelo menos três casos válidos. A análise é baseada em dados de número inteiro positivo. A opção de discretização categorizará automaticamente uma variável de valor fracionário, agrupando os seus valores em categorias com uma distribuição perto do "normal" e converterá automaticamente valores de variáveis de sequência em números inteiros positivos. É possível especificar outros esquemas de discretização.

Procedimentos relacionados. O ajuste de escala de todas as variáveis no nível numérico corresponde a análise de componentes principais padrão. Os recursos alternativos de plotagem estão disponíveis usando as variáveis transformadas em uma análise de componentes principais linear padrão. Se todas as variáveis tiverem níveis de ajuste de escala nominais múltiplos, a análise de componentes principais de dados categóricos é idêntica à análise de múltipla correspondência. Se os conjuntos de variáveis forem de interesse, a análise de correlação canônica categórica (não lineares) deve ser usada.

Para obter uma análise de componentes principais para dados categóricos

1. Nos menus, escolha:

Analisar > Redução de dimensão > Ajuste de escala ideal...

2. Selecione **Algumas variáveis não são nominais múltiplas**.
3. Selecione **Um conjunto**.
4. Clique em **Definir**.
5. Selecione pelo menos duas variáveis de análise e especifique o número de dimensões na solução.
6. Clique em **OK**.

É possível especificar opcionalmente variáveis complementares, que são ajustadas na solução encontrada ou rotular variáveis para os gráficos.

Definir escala e ponderação em CATPCA

É possível configurar o nível de ajuste de escala ideal para variáveis de análise e variáveis complementares. Por padrão, elas são escalas como splines monotônicas de segundo grau (ordinal) com dois nós internos. Além disso, é possível configurar a ponderação para variável de análise.

Ponderação variável. É possível escolher definir uma ponderação para cada variável. O valor especificado deve ser um número inteiro positivo. O valor padrão é 1.

Nível de ajuste de escala ideal. Também é possível selecionar o nível de ajuste de escala a ser usado para quantificar cada variável.

- **Ordinal spline.** A ordem das categorias da variável observada é preservada na variável idealmente escalada. Os pontos de categoria estarão em uma linha reta (vetor) até a origem. A transformação resultante é um polinomial monotônico suave em trechos do grau escolhido. As partes são especificadas pelo número especificado pelo usuário e pela colocação determinada pelo procedimento dos nós internos.
- **Nominal spline.** A única informação na variável observada que é preservada na variável idealmente escalada é o agrupamento de objetos em categorias. A ordem das categorias da variável observada não é preservada. Os pontos de categoria estarão em uma linha reta (vetor) até a origem. A transformação resultante é um polinomial não monotônico suave em trechos do grau escolhido. As partes são especificadas pelo número especificado pelo usuário e pela colocação determinada pelo procedimento dos nós internos.
- **Nominal múltiplo.** A única informação na variável observada que é preservada na variável idealmente escalada é o agrupamento de objetos em categorias. A ordem das categorias da variável observada não é preservada. Os pontos de categoria estarão no centroide dos objetos nas categorias particulares. *Múltiplo* indica que diferentes conjuntos de quantificações são obtidos para cada dimensão.
- **Ordinal.** A ordem das categorias da variável observada é preservada na variável escalada de modo ideal. Os pontos de categoria estarão em uma linha reta (vetor) até a origem. A transformação resultante se ajusta melhor que a transformação ordinal de spline, mas é menos suave.
- **Nominal.** A única informação na variável observada que é preservada na variável idealmente escalada é o agrupamento de objetos em categorias. A ordem das categorias da variável observada não é preservada. Os pontos de categoria estarão em uma linha reta (vetor) até a origem. A transformação resultante se ajusta melhor que a transformação nominal de spline, mas é menos suave.
- **Numérico.** As categorias são tratadas conforme ordenadas e igualmente espaçadas (nível de intervalo). A ordem das categorias e as distâncias iguais entre os números de categoria da variável observada são preservadas na variável idealmente escalada. Os pontos de categoria estarão em uma linha reta (vetor) até a origem. Quando todas as variáveis estão no mesmo nível numérico, a análise é análoga à análise de componentes principais padrão.

Discretização de análise de componentes principais para dados categóricos

A caixa de diálogo Discretização permite selecionar um método de gravar suas variáveis. As variáveis com valores fracionários são agrupadas em sete categorias (ou no número de valores distintos da variável, se esse número for menor que sete) com uma distribuição aproximadamente normal, a menos que seja especificado de outra forma. As variáveis de sequência são sempre convertidas em números inteiros positivos, designando indicadores de categoria de acordo com a ordem alfanumérica ascendente. A discretização para variáveis de sequência é aplicada a esses números inteiros. Outras variáveis são deixadas sozinhas por padrão. As variáveis discretizadas são, então, usadas na análise.

Método. Escolha entre agrupamento, ranqueamento e multiplicação.

- **Agrupamento.** Recodifique para um número especificado de categorias ou recodifique por intervalo.
- **Ranqueamento.** A variável é discretizada pelo pelos postos dos casos.
- **Multiplicação.** Os valores atuais da variável são padronizados, multiplicados por 10, arredondados e têm uma constante incluída, de tal modo que o menor valor discretizado é 1.

Agrupamento. Os pontos a seguir estão disponíveis ao discretizar variáveis por agrupamento:

- **Número de categorias.** Especifique o número de categorias e se os valores da variável devem seguir uma distribuição aproximadamente normal ou uniforme entre essas categorias.
- **Intervalos iguais.** As variáveis são recodificadas para categorias definidas por esses intervalos dimensionados igualmente. Deve-se especificar o comprimento dos intervalos.

Valores omissos da análise de componentes principais para dados categóricos

Use a caixa de diálogo Valores Omissos para escolher a estratégia para manipular valores omissos nas variáveis de análise e variáveis complementares.

Estratégia. Escolha excluir valores omissos (tratamento passivo), imputar valores omissos (tratamento ativo) ou exclua objetos com valores omissos (exclusão de listwise).

- **Exclua valores omissos para entrada de correlações após a quantificação.** Os objetos com valores omissos na variável selecionada não contribuem com a análise para esta variável. Se todas as variáveis recebem tratamento passivo, então os objetos com valores omissos em todas as variáveis são tratadas como complementar. Se as correlações são especificadas na caixa de diálogo Saída, (após análise) os valores omissos são imputados com a categoria mais frequente, ou modo, da variável para as correlações das variáveis originais. Para as correlações das variáveis escaladas idealmente, é possível escolher o método de entrada.
 - **Modo.** Substitua os valores omissos pelo modo da variável idealmente escalada.
 - **Categoria extra.** Substitua valores omissos pela quantificação de uma categoria extra. Esta configuração implica que os objetos com um valor omissos nesta variável são considerados para pertencer à mesma categoria (extra).
 - **Categoria aleatória.** Impute cada valor omissos em uma variável com o valor quantificado de um número de categorias aleatórias diferentes com base nas frequências marginais das categorias da variável.
- **Imputar valores omissos.** Os objetos com valores omissos na variável selecionada têm estes valores imputados. É possível escolher o método de imputação .
 - **Modo.** Substitua os valores omissos pela categoria mais frequente. Quando houver vários nós, aquele com o indicador de categoria menor será usado.
 - **Categoria extra.** Substitua os valores omissos com a mesma quantificação de uma categoria extra. Esta configuração implica que os objetos com um valor omissos nesta variável são considerados para pertencer à mesma categoria (extra).

- **Categoria aleatória.** Substitua cada valor omissos em uma variável pelo número de categorias aleatórias diferentes com base nas frequências marginais das categorias.
- **Exclua os objetos com valores omissos nesta variável.** Os objetos com valores omissos na variável selecionada são excluídos da análise. Esta estratégia não está disponível para variáveis complementares.

Opções de análise de componentes principais para dados categóricos

A caixa de diálogo Opções fornece controles para selecionar a configuração inicial, especificar critérios de interação e convergência, selecionar um método de normalização, escolher o método para rotular gráficos e especificar objetos complementares.

Objetos complementares. Especifique o número de caso do objeto ou o primeiro e o último número de caso de um intervalo de objetos que você deseja tornar complementar e, em seguida, clique em **Incluir**. Se um objeto for especificado como complementar, as ponderações de caso serão ignoradas para este objeto.

Método de Normalização. É possível especificar uma das cinco opções para normalizar os escores do objeto e as variáveis. Apenas um método de normalização pode ser usado em cada análise.

- **Variável Principal.** Esta opção otimiza a associação entre as variáveis. As coordenadas das variáveis no espaço objeto são os carregamentos de componente (correlações com componentes principais, como dimensões e escores do objeto). Este método é útil quando se está interessado principalmente na correlação entre as variáveis.
- **Objeto Principal.** Esta opção otimiza distâncias entre objetos. Este método é útil quando se está principalmente interessado em diferenças ou similaridades entre os objetos .
- **Simétrica.** Use esta opção de normalização se você estiver interessado principalmente na relação entre objetos e variáveis.
- **Independente.** Use esta opção de normalização se desejar examinar as distâncias entre objetos e correlações entre as variáveis separadamente.
- **Customizado.** É possível especificar qualquer valor real no intervalo encerrado $[-1, 1]$. Um valor de 1 é igual ao método Objeto Principal. Um valor de 0 é igual ao método Simétrico. Um valor de -1 é igual ao método Variável Principal. Ao especificar um valor maior que -1 e menor que 1, é possível difundir o autovalor sobre objetos e variáveis. Este método é útil para fazer um biplot ou triplot sob medida.

Crítérios. É possível especificar o número máximo de iterações que o procedimento pode percorrer em seus cálculos. Também é possível selecionar um valor de critério de convergência. O algoritmo para a iteração se a diferença no ajuste total entre as duas últimas iterações for menor que o valor de convergência ou se o número máximo de iterações for atingido.

Gráficos por rótulo. É possível especificar se as variáveis e rótulos de valor ou os nomes e valores das variáveis são usados nos gráficos. Também é possível especificar um comprimento máximo para rótulos.

Dimensões de gráfico. É possível controlar as dimensões que são exibidas na saída.

- **Exibir todas as dimensões na solução.** Todas as dimensões na solução são exibidas em uma matriz de gráfico de dispersão.
- **Restringir o número de dimensões.** As dimensões exibidas são restritas a pares plotados. Se você restringir as dimensões, deverá selecionar as dimensões mais baixas e mais altas a serem plotadas. A dimensão mais baixa pode ir de 1 ao número de dimensões na solução menos 1 e é plotada com relação às dimensões mais altas. O valor da dimensão mais alta pode ir de 2 ao número de dimensões na solução e indica a dimensão mais alta a ser usada na plotagem dos pares de dimensão. Esta especificação se aplica a todos os gráficos multidimensionais solicitados.

Rotação. É possível selecionar um método de rotação para obter resultados girados.

Nota: esses métodos de rotação não estão disponíveis se você selecionar **Executar autoinicialização** no diálogo **Bootstrap**.

- **Varimax**. Um método de rotação ortogonal que minimiza o número de variáveis que possuem altos carregamentos em cada componente. Este método simplifica a interpretação dos componentes.
- **Quartimax**. Um método de rotação que minimiza o número de componentes necessários para explicar cada variável. Este método simplifica a interpretação das variáveis observadas.
- **Equamax**. Um método de rotação é uma combinação do método Varimax, que simplifica os componentes, e do método Quartimax, que simplifica as variáveis. O número de variáveis que são altamente carregadas em um componente e o número de componentes necessários para explicar uma variável são minimizados.
- **Oblimin**. Um método de rotação oblíqua (não ortogonal). Quando delta for igual a 0, os componentes são mais oblíquos. Conforme delta se torna mais negativo, os componentes se tornam menos oblíquos. Os valores positivos permitem correlação de componente adicional. O valor de **Delta** deve ser menor ou igual a 0.8.
- **Promax**. Uma rotação oblíqua (não ortogonal), que permite que os componentes sejam correlacionados. Essa rotação pode ser calculada mais rapidamente que uma rotação Oblimin direta, sendo útil para grandes conjuntos de dados. A quantia de correlação (obliquidade) que é permitida é controlada pelo parâmetro kappa. O valor de **Kappa** deve ser maior ou igual a 1 e menor que 10.000.

Configuração. É possível ler dados de um arquivo que contém as coordenadas de uma configuração. A primeira variável no arquivo contém as coordenadas para a primeira dimensão. A segunda variável contém as coordenadas para a segunda dimensão, e assim por diante.

- **Inicial**. A configuração no arquivo que é especificado é usada como o ponto de início da análise.
- **Fixo**. A configuração no arquivo que é especificado é usada para ajuste nas variáveis. As variáveis que são ajustadas devem ser selecionadas como variáveis de análise, mas porque a configuração é corrigida, elas são tratadas como variáveis complementares (de modo que não precisam ser selecionadas como variáveis complementares).

Saída de análise de componentes principais para dados categóricos

A caixa de diálogo Saída controla a exibição de resultados.

Escores do objeto. Exibe os escores do objeto e possui as opções a seguir:

- **Incluir categorias de.** Exibe os indicadores de categoria das variáveis de análise selecionadas.
- **Rotular escores do objeto por.** A partir da lista de variáveis que são especificadas como variáveis de rotulagem, é possível selecionar uma para rotular os objetos.

Carregamentos de componente. Exibe os carregamentos de componente para todas as variáveis que não receberam múltiplos níveis de ajuste de escala nominal. É possível classificar os carregamentos de componente por tamanho.

Histórico de iteração. Para cada iteração, a variação contabilizada, a perda e o aumento na variação contabilizada são mostrados.

Correlações de variáveis originais. Mostra a matriz de correlações das variáveis originais e os autovalores desta matriz.

Correlações de variáveis transformadas. Mostra a matriz de correlações das variáveis transformadas (idealmente escalada) e os autovalores desta matriz.

Varição contabilizada. Exibe a quantia de variação contabilizada por coordenadas do centroide, coordenadas do vetor e total (coordenadas do centroide e do vetor combinadas) por variável e por dimensão.

Quantificações de categoria. Fornece as quantificações de categoria e as coordenadas para cada dimensão das variáveis que são selecionadas.

Estatísticas descritivas. Exibe frequências, número de valores omissos e modo das variáveis que são selecionados.

Salvamento da análise de componentes principais para dados categóricos

Os controles na caixa de diálogo Salvar salvam dados discretizados, escores do objeto, valores transformados e outros resultados no conjunto de dados ativo, em um novo conjunto de dados na sessão atual ou em um arquivo externo.

- Os conjuntos de dados estão disponíveis durante a sessão atual, mas não estão disponíveis nas sessões subsequentes, a menos que sejam salvos explicitamente como arquivos de dados. Os nomes dos conjuntos de dados devem aderir às regras de nomenclatura de variáveis.
- Os nomes dos arquivos ou os nomes do conjunto de dados devem ser diferentes para cada tipo de dado salvo.
- Se você salvar escores do objeto ou valores transformados no conjunto de dados ativo, será possível especificar o número de múltiplas dimensões nominais.
- As opções no grupo **Elipses de confiança do bootstrap** estão disponíveis somente se você selecionar **Executar autoinicialização** no diálogo Bootstrap.

Gráficos de objeto de análise de componentes principais para dados categóricos

A caixa de diálogo Gráficos do Objeto e da Variável permite especificar os tipos de gráficos que você deseja e as variáveis para as quais gráficos serão produzidos.

Pontos de objeto. Um gráfico dos pontos de objeto é exibido.

Objetos e variáveis (biplot). Os pontos de objetos são plotados com sua escolha das coordenadas de variável, carregamentos de componentes ou centroides da variável.

Objetos, carregamentos e centroides (triplot). Os pontos de objetos são plotados com os centroides das variáveis de nível de ajuste de escala nominal múltiplo e os carregamentos de componentes de outras variáveis.

Variáveis biplot e triplot. É possível escolher usar todas as variáveis para os biplots e triplots ou selecionar um subconjunto.

Rotulas objetos. É possível escolher ter objetos rotulados com as categorias das variáveis selecionadas (deve-se escolher valores indicadores de categoria ou rótulos de valores na caixa de diálogo Opções) ou com seus números de caso. Um gráfico é produzido por variável se a **Variável** for selecionada.

Gráficos de categoria de Análise de componentes principais para dados categóricos

A caixa de diálogo Gráficos de Categoria permite especificar os tipos de gráficos que você deseja e as variáveis para as quais os gráficos serão produzidos.

Gráficos de categoria. Para cada variável selecionada, um gráfico das coordenadas do centroide e vetor é plotado. Para as variáveis com vários níveis de ajuste de escala nominal, as categorias estão nos centroides dos objetos nas categorias particulares. Para todos os outros níveis de ajuste de escala, as categorias estão em um vetor que passa pela origem.

Juntar gráficos de categoria. Este é um gráfico único das coordenadas do centroide e vetor de cada variável selecionada.

Gráficos de transformação. Exibe um gráfico das qualificações da categoria ideal versus os indicadores de categoria. É possível especificar o número de dimensões para variáveis com vários níveis de ajuste de escala nominal. Também é possível escolher exibir gráficos de resíduos para cada variável selecionada.

Centroides projetados de. É possível escolher uma variável e projetar seus centroides em variáveis selecionadas. As variáveis com vários níveis de ajuste de escala nominal não podem ser selecionadas para o projeto ativo. Quando este gráfico é solicitado, uma tabela com as coordenadas dos centroides projetados também é exibida.

Gráficos de carregamento da análise de componentes principais para dados categóricos

A caixa de diálogo Gráficos de carregamento controla as variáveis que são incluídas no gráfico, a exibição de centroides no gráfico de carregamentos e a exibição de gráficos de variação representados.

Variação contabilizada. Para cada dimensão, é exibido um gráfico de variação contabilizada.

Exibição de carregamentos de componente Se selecionado, um gráfico dos carregamentos de componentes é exibido.

Variáveis de carregamento. É possível escolher usar todas as variáveis para o gráfico de carregamentos de componente ou selecionar um subconjunto.

Incluir centroides. As variáveis com níveis de ajuste de escala nominais múltiplos não têm carregamentos de componentes, mas é possível escolher incluir os centroides dessas variáveis no gráfico. É possível escolher usar todas as variáveis nominais ou selecionar um subconjunto.

Bootstrap de análise de componentes principais para dados categóricos

O diálogo Bootstrap especifica os parâmetros para análise de bootstrap.

Executar autoinicialização. Executa reamostragem de bootstrap.

- Se forem solicitados gráficos de carregamentos, categorias ou escores de componentes, gráficos extras serão exibidos. Esses gráficos exibem os pontos para a amostra de dados e as estimativas bootstrap. Os gráficos de transformação incluem regiões de confiança. Um gráfico para os autovalores também é exibido.
- Se uma solução bidirecional for especificada, os gráficos de elipse de confiança para os autovalores, os carregamentos de componentes, os pontos de categoria e os pontos de objetos serão exibidos.
- A reamostragem de bootstrap não estará disponível se você especificar um método de rotação no diálogo Opções.

É possível selecionar autoinicialização **Balanceada** ou **Não balanceada**.

Número de amostras. O número de amostras bootstrap que são usadas para calcular as estimativas bootstrap. O valor deve ser um número inteiro positivo.

Nível de confiança. O nível de confiança das estimativas bootstrap, expresso como uma porcentagem. O valor deve ser um número positivo menor que 100.

Método de correspondência. As duas alternativas são **Procrusto** e **Reflexão**.

Elipses de confiança. Controla a área de limite para elipses de confiança nos gráficos. Para áreas maiores que (> operador), o valor especificado, o número de elipses diminui conforme o valor do limite aumenta.

- As configurações neste grupo estão disponíveis apenas se o número de dimensões que são especificadas no diálogo principal for 2.
- Elipses de confiança para gráficos de carregamento estão disponíveis apenas se **Exibir carregamentos de componentes** estiver selecionado no diálogo Gráficos de Carregamento.
- Elipses de confiança para gráficos de objeto estão disponíveis apenas se **Pontos de objeto** estiver selecionado no diálogo Gráficos de Objetos e Variáveis.
- Elipses de confiança para gráficos de categoria estão disponíveis apenas se uma ou mais variáveis forem especificadas na lista **Gráficos de categoria** no diálogo Gráficos de Categoria.

Número de pontos de delineamento de elipse. O número dos pontos de gráfico usados para colher cada elipse de confiança. Valores maiores produzem elipses mais suaves. O valor deve ser um número inteiro positivo menor ou igual a 100.

Recursos adicionais do comando CATPCA

É possível customizar sua análise de componentes principais para dados categóricos se você colar suas seleções em uma janela de sintaxe e editar a sintaxe de comando CATPCA resultante. O idioma da sintaxe de comando também permite:

- Especifique nomes de raiz para as variáveis transformadas, escores de objeto e aproximações ao salvá-las para o conjunto de dados ativo (com o subcomando SAVE).
- Especifique um comprimento máximo para rótulos para cada gráfico separadamente (com o subcomando PLOT).
- Especifique uma lista de variáveis separada para gráficos de resíduos (com o subcomando PLOT).

Consulte a *Referência da sintaxe de comando* para obter informações de sintaxe completa.

Análise de correlação canônica não linear (GERAIS)

A análise de correlação canônica não linear corresponde à análise de correlação canônica categórica com ajuste de escala ideal. O propósito deste procedimento é determinar o quão semelhantes os conjuntos de variáveis categóricas são entre si. A análise de correlação canônica não linear também é conhecida pelo acrônimo OVERALS.

A análise de correlação canônica padrão é uma extensão de múltipla regressão, em que o segundo conjunto não contém uma variável de resposta única, em vez disso, contém variáveis de respostas diversas. O objetivo é explicar o máximo possível da variância nos relacionamentos entre os dois conjuntos de variáveis numéricas em um espaço dimensional baixo. Inicialmente, as variáveis em cada conjunto são linearmente combinadas de tal forma que as combinações lineares têm uma correlação máxima. Dadas estas combinações, as combinações lineares subsequentes são determinadas como não correlacionadas com as combinações anteriores e têm a maior correlação possível.

A abordagem de ajuste de escala ideal expande a análise padrão de três formas cruciais. Primeiro, OVERALS permite mais de dois conjuntos de variáveis. Segundo, as variáveis podem ser escaladas como nominal, ordinal ou numérica. Como resultado, os relacionamentos não lineares entre as variáveis podem ser analisados. Finalmente, em vez de maximizar correlações entre os conjuntos de variáveis, os conjuntos são comparados a um conjunto de compromissos desconhecido que é definido pelos escores do objeto.

Exemplo. A análise de correlação canônica categórica com ajuste de escala ideal poderia ser usada para apresentar graficamente o relacionamento entre um conjunto de variáveis que contém categoria de trabalho e anos de escolaridade e outro conjunto de variáveis contendo a região de residência e o sexo. É possível descobrir que anos de escolaridade e a região de residência discriminam melhor do que as variáveis restantes. Também é possível descobrir que anos de escolaridade discriminam melhor a primeira dimensão.

Estatísticas e gráficos. Frequências, centroides, histórico de iteração, escores do objeto, quantificações de categoria, ponderações, carregamentos de componente, ajuste único e múltiplo, gráficos de escores do objeto, gráficos de coordenadas de categoria, gráficos de carregamentos de componentes, gráficos de centroides de categoria, gráficos de transformação.

Considerações de dados de análise de correlação canônica não linear

Dados. Use números inteiros para codificar variáveis (nível de ajuste de escala nominal ou ordinal). Para minimizar a saída, use números inteiros consecutivos com 1 para o código de cada variável. As variáveis que são escaladas no nível numérico não devem ser recodificadas para números inteiros consecutivos. Para minimizar a saída, para cada variável que é escalada no nível numérico, subtraia o menor valor observado de cada valor e inclua 1. Os valores fracionais são truncados após o decimal.

Suposições. As variáveis podem ser classificadas em dois ou mais conjuntos. As variáveis na análise são escaladas como nominais múltiplas, nominais únicas, ordinal ou numérica. O número máximo de dimensões que são usadas no procedimento depende do nível de ajuste de escala ideal das variáveis. Se todas as variáveis são especificadas como ordinal, nominal única, ou numérica, o número máximo de dimensões é o menor dos dois valores a seguir: o número de observações menos 1 ou o número total de variáveis. No entanto, se apenas dois conjuntos de variáveis são definidos, o número máximo de dimensões é o número de variáveis no conjunto menor. Se algumas variáveis são nominais múltiplas, o número máximo de dimensões é o número total de categorias nominais múltiplas mais o número de variáveis nominais não múltiplas menos o número de variáveis nominais múltiplas. Por exemplo, se a análise envolver cinco variáveis, um das quais é nominal múltipla com quatro categorias, o número máximo de dimensões é $(4 + 4 - 1)$ ou 7. Se você especificar um número maior que o máximo, o valor máximo é usado.

Procedimentos relacionados. Se cada conjunto contém uma variável, a análise de correlação canônica não linear é equivalente à análise de componentes principais com ajuste de escala ideal. Se cada uma dessas variáveis for nominal múltipla, a análise corresponde à análise de múltipla correspondência. Se dois conjuntos de variáveis estão envolvidos, e um dos conjuntos contém apenas uma variável, a análise é idêntica à regressão categórica com ajuste de escala ideal.

Para obter uma análise de correlação canônica não linear

1. Nos menus, escolha:

Analisar > Redução de dimensão > Ajuste de escala ideal...

2. Selecione **Todas as variáveis nominais múltiplas** ou **Algumas variáveis nominais não múltiplas**.

3. Selecione **Conjuntos múltiplos**.

4. Clique em **Definir**.

5. Defina, pelo menos, dois conjuntos de variáveis. Selecione a(s) variável(is) que você deseja incluir no primeiro conjunto. Para mover para o próximo conjunto, clique em **Avançar** e selecione as variáveis que você deseja incluir no segundo conjunto. É possível incluir conjuntos adicionais. Clique em **Anterior** para retornar ao conjunto de variáveis definido anteriormente.

6. Defina o intervalo de valores e a escala de medida (nível de ajuste de escala ideal) para cada variável selecionada.

7. Clique em **OK**.

8. Opcionalmente:

- Selecione uma ou mais variáveis para fornecer rótulos de pontos para gráficos de escore do objeto. Cada variável produz um gráfico separado, com os pontos rotulados pelos valores desta variável. Deve-se definir um intervalo para cada uma dessas variáveis de rótulo de gráfico. Ao usar a caixa de diálogo, uma variável única não pode ser usada na análise e como uma variável de rotulagem. Se você quiser rotular o gráfico de escores do objeto com uma variável que é usada na análise, use o recurso Calcular (disponível a partir do menu Transformar) para criar uma cópia dessa variável. Use a nova variável para rotular o gráfico. Como alternativa, a sintaxe de comando pode ser usada.

- Especifique o número de dimensões que você deseja na solução. Em geral, escolha algumas dimensões, conforme necessário, para explicar a maior parte da variação. Se a análise envolve mais do que duas dimensões, gráficos tridimensionais das três primeiras dimensões são produzidos. Outras dimensões podem ser exibidas ao editar o gráfico.

Definir intervalo e escala

Deve-se definir um intervalo para cada variável. O valor máximo que é especificado deve ser um número inteiro. Os valores de dados fracionais são truncados na análise. Um valor de categoria que está fora do intervalo especificado é ignorado na análise. Para minimizar a saída, use o recurso Recodificação automática (disponível no menu Transformar) para criar categorias consecutivas iniciando com 1 para variáveis que são tratadas como nominal ou ordinal. A recodificação de números inteiros consecutivos não é recomendada para variáveis que são escalonadas no nível numérico. Para minimizar a saída para variáveis que são tratadas como numéricas, para cada variável, subtraia o valor mínimo de cada valor e inclua o 1.

Deve-se também selecionar o ajuste de escala para ser usado para quantificar cada variável.

- **Ordinal.** A ordem das categorias da variável observada é preservada na variável quantificada.
- **Nominal única.** Na variável quantificada, os objetos na mesma categoria recebem a mesma pontuação.
- **Multinomial.** As quantificações podem ser diferentes para cada dimensão.
- **Númerico discreto.** As categorias são tratadas conforme ordenadas e igualmente espaçadas. As diferenças entre os números de categorias e a ordem das categorias da variável observada são preservadas na variável quantificada.

Definir intervalo

Deve-se definir um intervalo para cada variável. O valor máximo que é especificado deve ser um número inteiro. Os valores de dados fracionais são truncados na análise. Um valor de categoria que está fora do intervalo especificado é ignorado na análise. Para minimizar a saída, use o recurso Recodificação Automática (disponível no menu Transformar) para criar categorias consecutivas iniciando com 1.

Deve-se também definir um intervalo para cada variável que é usada para rotular os gráficos de escores do objeto. No entanto, rótulos para categorias com valores de dados que estão fora do intervalo definido para a variável aparecem nos gráficos.

Opções de análise de correlação canônica não linear

A caixa de diálogo Opções permite selecionar estatísticas e gráficos opcionais, salvar escores de objetos como novas variáveis no conjunto de dados ativo, especificar critérios de iteração e convergência e especificar uma configuração inicial para a análise.

Exibição. As estatísticas disponíveis incluem frequências marginais (contagens), centroides, histórico de iteração, ponderações e carregamentos de componente, quantificações de categoria, escores do objeto e estatísticas de ajuste únicas e múltiplas.

- *Centroides.* Quantificações de categoria e as médias projetadas e reais dos escores do objeto para os objetos (casos) incluídos em cada conjunto com relação aos casos que pertencerem à mesma categoria da variável.
- *Carregamentos de componente e ponderações.* Os coeficientes de regressão em cada dimensão para cada variável quantificada em um conjunto, em que os escores de objeto são regredidos nas variáveis quantificadas, e a projeção da variável quantificada no espaço de objetos. Fornece uma indicação da contribuição que cada variável faz para a dimensão dentro de cada conjunto.
- *Ajuste único e Ajuste múltiplo.* Medidas da qualidade do ajuste de coordenadas de categoria/quantificações de categoria única e múltiplas com relação aos objetos.
- *Quantificações de categoria.* Valores de escala ideais designados para as categorias de uma variável.
- *Escores do objeto.* Escore ideal designado a um objeto (caso) em uma dimensão específica.

Gráfico. É possível produzir gráficos de coordenadas de categoria, escores do objeto, carregamentos de componente, centroides de categoria e transformações.

Salvar escores do objeto. É possível salvar os escores do objeto como novas variáveis no conjunto de dados ativo. Os escores do objeto são salvos para o número de dimensões que são especificados na caixa de diálogo principal.

Usar configuração inicial aleatória. Uma configuração inicial aleatória deve ser usada se algumas ou todas as variáveis forem nominais únicas. Se esta opção não for selecionada, uma configuração inicial aninhada será usada.

Critérios. É possível especificar o número máximo de iterações que a análise de correlação canônica não linear pode percorrer em seus cálculos. Também é possível selecionar um valor de critério de convergência. A análise interrompe a iteração se a diferença no ajuste total entre as últimas duas iterações for menor que o valor de convergência ou se o número máximo de iterações for atingido.

Recursos adicionais de comando OVERALS

É possível customizar sua análise de correlação canônica não linear se você colar suas seleções em uma janela de sintaxe e editar a sintaxe de comando OVERALS resultante. O idioma da sintaxe de comando também permite:

- Especificar os pares de dimensão a serem plotados, em vez de plotar todas as dimensões extraídas (usando a palavra-chave NDIM no subcomando PLOT).
- Especificar o número de caracteres de rótulo de valor que são usados para rotular pontos nos gráficos (com o subcomando PLOT).
- Designar mais que cinco variáveis como variáveis de rotulagem para gráficos de escores do objeto (com o subcomando PLOT).
- Selecionar variáveis que são usadas na análise como variáveis de rotulagem para os gráficos de escores do objeto (com o subcomando PLOT).
- Selecionar variáveis para fornecer rótulos de ponto para o gráfico de escore de quantificação (com o subcomando PLOT).
- Especificar o número de casos a serem incluídos na análise se você não quiser usar todos os casos no conjunto de dados ativo (com o subcomando NOBSERVATIONS).
- Especificar nomes de raízes para variáveis criadas ao salvar escores do objeto (com o subcomando SAVE).
- Especificar o número de dimensões a ser salvo, em vez de salvar todas as dimensões extraídas (com o subcomando SAVE).
- Gravar as quantificações de categoria em um arquivo matriz (usando o subcomando MATRIX).
- Produzir gráficos de baixa resolução que podem ser mais fáceis de ler que os gráficos de alta resolução comuns (usando o comando SET).
- Produzir gráficos de centroide e transformação somente para variáveis especificadas (com o subcomando PLOT).

Consulte a *Referência da sintaxe de comando* para obter informações de sintaxe completa.

Análise de correspondências

Um dos objetivos da análise de correspondência é descrever os relacionamentos entre duas variáveis nominais em uma tabela de correspondência em um espaço de baixa dimensão, ao mesmo tempo em que descreve os relacionamentos entre as categorias para cada variável. Para cada variável, as distâncias entre os pontos de categoria em um gráfico refletem os relacionamentos entre as categorias com categorias similares plotadas próximas umas das outras. A projeção de pontos para uma variável no vetor a partir da origem para um ponto de categoria para a outra variável descreve o relacionamento entre as variáveis.

Uma análise de tabelas de contingência frequentemente inclui examinar perfis de linha e coluna e testar a independência por meio de estatística qui-quadrado. No entanto, o número de perfis pode ser muito grande e o teste qui-quadrado não revela a estrutura de dependência. O procedimento de Tabulações cruzadas oferece várias medidas de associação e testes de associação, mas não é possível representar graficamente qualquer relacionamento entre as variáveis.

A análise fatorial é uma técnica padrão para descrever relacionamentos entre variáveis em um espaço de baixa dimensão. No entanto, a análise fatorial requer dados do intervalo, e o número de observações deve ser cinco vezes o número de variáveis. A análise de correspondência, por outro lado, assume as variáveis nominais e pode descrever os relacionamentos entre categorias de cada variável, bem como o relacionamento entre as variáveis. Além disso, a análise de correspondência pode ser usada para analisar qualquer tabela de medidas de correspondência positivas.

Exemplo. A análise de correspondência poderia ser usada para apresentar graficamente a relação entre a categoria de equipe e os hábitos de fumar. É possível descobrir que, em relação ao fumo, gerentes juniores diferem de secretários, mas secretários não diferem dos gerentes seniores. Também é possível descobrir que o tabagismo pesado está associado aos gerentes juniores, considerando que o tabagismo leve está associado aos secretários.

Estatísticas e gráficos. Medidas de correspondência, perfis de linha e coluna, valores singulares, escores de linhas e colunas, inércia, massa, estatísticas de confiança de escore de linha e coluna, estatísticas de confiança de valor singular, gráficos de transformação, gráficos de pontos de linha, gráficos de pontos de coluna e biplots.

Considerações de dados de análise de correspondência

Dados. Variáveis categóricas a serem analisadas são escaladas nominalmente. Para dados agregados ou para uma medida de correspondência diferente de frequências, use uma variável de ponderação com valores de similaridades positivos. Como alternativa, para os dados da tabela, use a sintaxe para ler a tabela.

Suposições. O número máximo de dimensões usadas no procedimento depende do número de categorias de linhas e colunas ativas e o número de restrições de igualdade. Se nenhuma restrição de igualdade é usada e todas as categorias estão ativas, a dimensionalidade máxima é aquela que é menor que o número de categorias para a variável com menos categorias. Por exemplo, se uma variável possuir cinco categorias e outra possuir quatro, o número máximo de dimensões será três. As categorias complementares não são ativas. Por exemplo, se uma variável possuir cinco categorias, das quais duas são complementares, e a outra possuir quatro, o número máximo de dimensões será duas. Trate todos os conjuntos de categorias que são restritas para serem iguais como uma categoria. Por exemplo, se uma variável possuir cinco categorias, três das quais restritas para serem iguais, esta variável deve ser tratada como tendo três categorias ao determinar a dimensionalidade máxima. Duas das categorias são irrestritas, e a terceira categoria corresponde às três categorias restritas. Se você especificar um número de dimensões maior que o máximo, o valor máximo será usado.

Procedimentos relacionados. Se mais de duas variáveis forem envolvidas, use a análise de múltipla correspondência. Se as variáveis devem ser escaladas ordinalmente, use a análise de componentes principais para dados categóricos.

Para obter uma análise de correspondência

1. Nos menus, escolha:
Analisar > Redução de dimensão > Análise de correspondência...
2. Selecione uma variável de linha.
3. Selecione uma variável da coluna.
4. Defina os intervalos para as variáveis.

5. Clique em OK.

Definir intervalo de linha na análise de correspondência

Deve-se definir um intervalo para a variável de linha. Os valores mínimo e máximo especificados devem ser números inteiros. Os valores de dados fracionais são truncados na análise. Um valor de categoria que está fora do intervalo especificado é ignorado na análise.

Todas as categorias são inicialmente sem restrição e ativas. É possível restringir categorias de linhas para outras categorias de linhas iguais, ou é possível definir uma categoria de linha como complementar.

- **As categorias devem ser iguais.** As categorias devem ter escores iguais. Use restrições de igualdade se a ordem obtida para as categorias for indesejada ou contraintuitiva. O número máximo de categorias de linhas que podem ser restringidas a serem iguais é o número total de categorias de linhas ativas menos 1. Para impor restrições de igualdade diferentes em conjuntos de categorias, use a sintaxe. Por exemplo, use a sintaxe para restringir categorias 1 e 2 a serem iguais e as categorias 3 e 4 a serem iguais.
- **A categoria é complementar.** As categorias complementares não influenciam a análise, mas são representadas no espaço definido pelas categorias ativas. Categorias complementares executam sem regra na definição das dimensões. O número máximo das categorias de linhas complementares é o número total das categorias de linhas menos 2.

Definir intervalo da coluna em análise de correspondência

Deve-se definir um intervalo para a variável da coluna. Os valores mínimo e máximo especificados devem ser números inteiros. Os valores de dados fracionais são truncados na análise. Um valor de categoria que está fora do intervalo especificado é ignorado na análise.

Todas as categorias são inicialmente sem restrição e ativas. É possível restringir categorias de coluna para outras categorias de colunas iguais, ou é possível definir uma categoria de coluna como complementar.

- **As categorias devem ser iguais.** As categorias devem ter escores iguais. Use restrições de igualdade se a ordem obtida para as categorias for indesejada ou contraintuitiva. O número máximo de categorias de coluna que podem ser restringidas a serem iguais é o número total de categorias de colunas ativas menos 1. Para impor restrições de igualdade diferentes nos conjuntos de categorias, use a sintaxe. Por exemplo, use a sintaxe para restringir categorias 1 e 2 a serem iguais e as categorias 3 e 4 a serem iguais.
- **A categoria é complementar.** As categorias complementares não influenciam a análise, mas são representadas no espaço definido pelas categorias ativas. Categorias complementares executam sem regra na definição das dimensões. O número máximo das categorias de colunas complementares é o número total das categorias de colunas menos 2.

Modelo de análise de correspondência

A caixa de diálogo Modelo permite especificar o número de dimensões, a medida de distância, o método de padronização e o método de normalização.

Dimensões na solução. Especifique o número de dimensões. Em geral, escolha algumas dimensões, conforme necessário, para explicar a maior parte da variação. O número máximo de dimensões depende do número de categorias ativas usadas na análise e nas restrições de igualdade. O número máximo de dimensões é o menor de:

- O número de categorias de linhas ativas menos o número de categorias de linhas restritas a serem iguais, mais o número de conjuntos de categorias de linhas restritas.
- O número de categorias de colunas ativas menos o número de categorias de colunas restritas a serem iguais, mais o número de conjuntos de categorias de colunas restritas.

Medida de distância. É possível selecionar a medida de distância entre as linhas e colunas da tabela de correspondência. Escolha uma das alternativas a seguir:

- **Qui-quadrado.** Use uma distância de perfil ponderada, em que o peso é a massa das linhas ou colunas. Esta medida é requerida para análises de correspondência padrão.
- **Euclidiana.** Use a raiz quadrada da soma das diferenças quadráticas entre pares de linhas e pares de colunas.

Método de padronização. Escolha uma das alternativas a seguir:

- **As médias de linha e coluna são removidas.** Tanto linhas como colunas são centralizadas. Este método é requerido para análise de correspondência padrão.
- **As médias de linhas são removidas.** Apenas as linhas são centralizadas.
- **As médias de coluna são removidas.** Apenas as colunas são centralizadas.
- **Os totais de linhas são equalizados e as médias são removidas.** Antes de centralizar as linhas, as margens de linha são equalizadas.
- **Os totais de coluna são equalizados e as médias são removidas.** Antes de centralizar as colunas, as margens da coluna são equalizadas.

Método de Normalização. Escolha uma das alternativas a seguir:

- **Simétrica.** Para cada dimensão, os escores da linha são a média ponderada dos escores da coluna, divididos pelo valor singular correspondente, e os escores da coluna são a média ponderada dos escores da linha, divididos pelo valor singular correspondente. Use esse método se desejar examinar as diferenças ou similaridades entre as categorias das duas variáveis.
- **Principal.** As distâncias entre os pontos de linha e os pontos de coluna são aproximações das distâncias na tabela de correspondência, de acordo com a medida de distância selecionada. Utilize este método se desejar examinar as diferenças entre as categorias de uma ou de ambas as variáveis ao invés de diferenças entre as duas variáveis.
- **Linha principal.** As distâncias entre os pontos de linha são aproximações das distâncias na tabela de correspondência, de acordo com a medida de distância selecionada. Os escores da linha são a média ponderada dos escores da coluna. Utilize este método se desejar examinar diferenças ou similaridades entre as categorias da variável de linha.
- **Coluna principal.** As distâncias entre os pontos de coluna são aproximações das distâncias na tabela de correspondência, de acordo com a medida de distância selecionada. Os escores da coluna são a média ponderada dos escores da linha. Use este método se desejar examinar diferenças ou similaridades entre as categorias da variável da coluna.
- **Customizado.** Deve-se especificar um valor entre -1 e 1 . Um valor de -1 corresponde à coluna principal. Um valor de 1 corresponde à linha principal. Um valor de 0 corresponde ao simétrico. Todos os outros valores difundem a inércia sobre os escores da linha e da coluna para graus de variação. Este método é útil para fazer biplots sob medida.

Estatísticas de análise de correspondência

A caixa de diálogo Estatísticas permite especificar a saída numérica produzida.

Tabela de correspondência. Uma tabulação cruzada das variáveis de entrada com totais marginais de linha e coluna.

Visão geral dos pontos da linha. Para cada categoria de linha, as pontuações, massa, inércia, contribuição para a inércia da dimensão e a contribuição da dimensão para a inércia do ponto.

Visão geral dos pontos da coluna. Para cada categoria de coluna, as pontuações, massa, inércia, contribuição para a inércia da dimensão e a contribuição da dimensão para a inércia do ponto.

Perfis de linha. Para cada categoria de linha, a distribuição entre as categorias da variável da coluna.

Perfis de coluna. Para cada categoria de coluna, a distribuição entre as categorias da variável de linha.

Permutações da tabela de correspondência. A tabela de correspondência reorganizada, de tal modo que as linhas e colunas estão por ordem crescente, de acordo com a as pontuações na primeira dimensão. Opcionalmente, é possível especificar o número máximo de dimensões para as quais as tabelas permutadas serão produzidas. Uma tabela permutada para cada dimensão, de 1 ao número especificado, é produzida.

Estatísticas de confiança para pontos de linha. Inclui o desvio padrão e as correlações para todos os pontos de linha não complementares.

Estatísticas de confiança para pontos de coluna. Inclui o desvio padrão e as correlações para todos os pontos de coluna não complementares.

Gráficos de análise de correspondência

A caixa de diálogo Gráficos permite especificar quais gráficos são produzidos.

Gráficos de dispersão. Produz uma matriz de todos os gráficos emparelhados das dimensões. Os gráficos de dispersão incluem:

- **Biplot.** Produz uma matriz de gráficos de conjunto dos pontos de linha e coluna. Se a normalização principal for selecionada, o biplot não fica disponível.
- **Pontos de linha.** Produz uma matriz de gráficos dos pontos de linha.
- **Pontos de coluna.** Produz uma matriz de gráficos dos pontos de coluna.

Opcionalmente, é possível especificar quantos caracteres de rótulo de valor usar ao rotular os pontos. Este valor deve ser um número inteiro não negativo menor ou igual a 20.

Gráficos de linha. Produz um gráfico para cada dimensão da variável selecionada. Os gráficos de linha disponíveis incluem:

- **Categorias de linha transformadas.** Produz um gráfico dos valores de categoria de linha originais com relação aos seus escores de linha correspondentes.
- **Categorias de coluna transformadas.** Produz um gráfico dos valores de categoria de coluna originais com relação aos seus escores de coluna correspondentes.

Opcionalmente, é possível especificar quantos caracteres de rótulo de valor usar ao rotular os eixos de categoria. Este valor deve ser um número inteiro não negativo menor ou igual a 20.

Dimensões de gráfico. Permite controlar as dimensões exibidas na saída.

- **Exibir todas as dimensões na solução.** Todas as dimensões na solução são exibidas em uma matriz de gráfico de dispersão.
- **Restringir o número de dimensões.** As dimensões exibidas são restritas a pares plotados. Se você restringir as dimensões, deverá selecionar as dimensões mais baixas e mais altas a serem plotadas. A dimensão mais baixa pode ir de 1 ao número de dimensões na solução menos 1 e é plotada com relação às dimensões mais altas. O valor da dimensão mais alta pode ir de 2 ao número de dimensões na solução e indica a dimensão mais alta a ser usada na plotagem dos pares de dimensão. Esta especificação se aplica a todos os gráficos multidimensionais solicitados.

Recursos adicionais do comando CORRESPONDENCE

É possível customizar sua análise de correspondência se você colar suas seleções em uma janela de sintaxe e editar a sintaxe de comando CORRESPONDENCE resultante. O idioma da sintaxe de comando também permite:

- Especifique os dados da tabela como entrada, em vez de usar dados casewise (usando o subcomando TABLE = ALL).
- Especifique o número de caracteres de rótulo de valor usados para rotular os pontos para cada tipo de matriz de gráfico de dispersão ou matriz biplot (com o subcomando PLOT).

- Especifique o número de caracteres de rótulo de valor usados para rotular os pontos para cada tipo de gráfico de linha (com o subcomando PLOT).
- Grave uma matriz de escores de linha e coluna em um arquivo de dados de matriz (com o subcomando OUTFILE).
- Grave uma matriz de estatísticas de confiança (variâncias e covariâncias) para os valores singulares e os escores em um arquivo de dados de matriz (com o subcomando OUTFILE).
- Especifique vários conjuntos de categorias para serem iguais (com o subcomando EQUAL).

Consulte a *Referência da sintaxe de comando* para obter informações de sintaxe completa.

Análise de correspondências múltipla

A Análise de Múltipla Correspondência quantifica dados nominais (categóricos) designando valores numéricos para os casos (objetos) e categorias, de modo que os objetos dentro da mesma categoria estão próximos e objetos em diferentes categorias estão distantes. Cada objeto fica tão perto quanto possível dos pontos de categoria das categorias que são aplicados ao objecto. Dessa forma, as categorias dividem os objetos em subgrupos homogêneos. As variáveis são consideradas homogêneas quando classificam os objetos nas mesmas categorias para os mesmos subgrupos.

Exemplo. A análise de múltipla correspondência poderia ser usada para apresentar graficamente a relação entre a categoria de cargo, classificação minoritária e sexo. É possível descobrir que a classificação minoritária e o sexo são discriminados entre pessoas, mas a categoria de cargo não. Também é possível descobrir que as categorias Latino e Afro-americano são semelhantes uma a outra.

Estatísticas e gráficos. Escores do objeto, medidas de discriminação, histórico de iteração, correlações de variáveis originais e transformadas, quantificações de categoria, estatísticas descritivas, gráficos de pontos de objeto, biplots, gráficos de categoria, gráficos de categoria de conjunto, gráficos de transformação e gráficos de medidas de discriminação.

Considerações de dados de análise de múltipla correspondência

Dados. Os valores de variável de sequência de caracteres são sempre convertidos em números inteiros positivos pela ordem alfanumérica ascendente. Valores omissos definidos pelo usuário, valores omissos do sistema e valores menores que 1 são considerados omissos; é possível recodificar ou incluir uma constante para variáveis com valores inferiores a 1 para torná-las não omissas.

Suposições. Todas as variáveis têm o nível de ajuste de escala nominal. Os dados devem conter pelo menos três casos válidos. A análise é baseada em dados de número inteiro positivo. A opção de discretização categorizará automaticamente uma variável de valor fracionário, agrupando os seus valores em categorias com uma distribuição perto do normal e converterá automaticamente valores de variáveis de sequência em números inteiros positivos. É possível especificar outros esquemas de discretização.

Procedimentos relacionados. Para duas variáveis, a Análise de múltipla correspondência é análoga à Análise de correspondência. Se você acredita que as variáveis possuem propriedades ordinais ou numéricas, a Análise de componentes principais para dados categóricos deve ser usada. Se os conjuntos de variáveis forem de interesse, a Análise de correlação canônica não linear deve ser usada.

Para obter uma Análise de múltipla correspondência

1. Nos menus, escolha:
 - Analisar > Redução de dimensão > Ajuste de escala ideal...**
2. Selecione **Todas as variáveis nominais múltiplas**.
3. Selecione **Um conjunto**.
4. Clique em **Definir**.
5. Selecione pelo menos duas variáveis de análise e especifique o número de dimensões na solução.

6. Clique em OK.

É possível especificar opcionalmente variáveis complementares, que são ajustadas na solução encontrada ou rotular variáveis para os gráficos.

Definir ponderação variável na análise de múltipla correspondência

É possível configurar a ponderação para variáveis de análise.

Ponderação variável. É possível escolher definir uma ponderação para cada variável. O valor especificado deve ser um número inteiro positivo. O valor padrão é 1.

Discretização de análise de correspondências múltipla

A caixa de diálogo Discretização permite selecionar um método de gravar suas variáveis. As variáveis de valor fracionário são agrupadas em sete categorias (ou no número de valores distintos da variável, se este número for menor que sete) com uma distribuição aproximadamente normal, a menos que seja especificado de outra forma. As variáveis de sequência são sempre convertidas em números inteiros positivos, designando indicadores de categoria de acordo com a ordem alfanumérica ascendente. A discretização para variáveis de sequência é aplicada a esses números inteiros. Outras variáveis são deixadas sozinhas por padrão. As variáveis discretizadas são, então, usadas na análise.

Método. Escolha entre agrupamento, ranqueamento e multiplicação.

- **Agrupamento.** Recodifique para um número especificado de categorias ou recodifique por intervalo.
- **Ranqueamento.** A variável é discretizada pelo pelos postos dos casos.
- **Multiplicação.** Os valores atuais da variável são padronizados, multiplicados por 10, arredondados e têm uma constante incluída, de modo que o menor valor discretizado é 1.

Agrupamento. Os pontos a seguir estão disponíveis ao discretizar variáveis por agrupamento:

- **Número de categorias.** Especifique o número de categorias e se os valores da variável devem seguir uma distribuição aproximadamente normal ou uniforme entre essas categorias.
- **Intervalos iguais.** As variáveis são recodificadas para categorias definidas por esses intervalos dimensionados igualmente. Deve-se especificar o comprimento dos intervalos.

Valores omissos da Análise de múltipla correspondência

A caixa de diálogo Valores omissos permite escolher a estratégia para tratamento de valores omissos em variáveis de análise e variáveis complementares.

Estratégia de valor omissos. Escolha excluir valores omissos (tratamento passivo), imputar valores omissos (tratamento ativo) ou exclua objetos com valores omissos (exclusão de listwise).

- **Exclua valores omissos para entrada de correlações após a quantificação.** Os objetos com valores omissos na variável selecionada não contribuem com a análise para esta variável. Se todas as variáveis recebem tratamento passivo, então os objetos com valores omissos em todas as variáveis são tratadas como complementar. Se as correlações são especificadas na caixa de diálogo Saída, (após análise) os valores omissos são imputados com a categoria mais frequente, ou modo, da variável para as correlações das variáveis originais. Para as correlações das variáveis escaladas idealmente, é possível escolher o método de entrada. Selecione **Modo** para substituir valores omissos com o modo da variável idealmente escalada. Selecione **Categoria extra** para substituir valores omissos com a quantificação de uma categoria extra. Isto implica que objetos com um valor omissos nesta variável são considerados para pertencer à mesma categoria (extra).
- **Imputar valores omissos.** Os objetos com valores omissos na variável selecionada têm estes valores imputados. É possível escolher o método de imputação. Selecione **Modo** para substituir valores omissos com a categoria mais frequente. Quando houver vários nós, aquele com o indicador de categoria menor será usado. Selecione **Categoria extra** para substituir valores omissos pela mesma

quantificação de uma categoria extra. Isto implica que objetos com um valor omissos nesta variável são considerados para pertencer à mesma categoria (extra).

- **Exclua os objetos com valores omissos nesta variável.** Os objetos com valores omissos na variável selecionada são excluídos da análise. Esta estratégia não está disponível para variáveis complementares.

Opções de análise de múltipla correspondência

A caixa de diálogo Opções permite selecionar a configuração inicial, especificar critérios de interação e convergência, selecionar um método de normalização, escolher o método para rotular gráficos e especificar objetos complementares.

Objetos complementares. Especifique o número de caso do objeto (ou o primeiro e o último número de caso de um intervalo de objetos) que você deseja tornar complementar e, em seguida, clique em **Incluir**. Continue até ter especificado todos os objetos complementares. Se um objeto for especificado como complementar, as ponderações de caso serão ignoradas para este objeto.

Método de Normalização. É possível especificar uma das cinco opções para normalizar os escores do objeto e as variáveis. Apenas um método de normalização pode ser usado em uma determinada análise.

- **Variável Principal.** Esta opção otimiza a associação entre as variáveis. As coordenadas das variáveis no espaço objeto são o os carregamentos de componente (correlações com componentes principais, como dimensões e escores do objeto). Este método é útil quando se está interessado principalmente na correlação entre as variáveis.
- **Objeto Principal.** Esta opção otimiza distâncias entre objetos. Este método é útil quando se está principalmente interessado em diferenças ou similaridades entre os objetos.
- **Simétrica.** Use esta opção de normalização se você estiver interessado principalmente na relação entre objetos e variáveis.
- **Independente.** Use esta opção de normalização se desejar examinar as distâncias entre objetos e correlações entre as variáveis separadamente.
- **Customizado.** É possível especificar qualquer valor real no intervalo encerrado $[-1, 1]$. Um valor de 1 é igual ao método Objetivo Principal, um valor de 0 é igual ao método Simétrico e um valor de -1 é igual ao método Variável Principal. Ao especificar um valor maior que -1 e menor que 1, é possível difundir o autovalor sobre objetos e variáveis. Este método é útil para fazer um biplot ou triplot sob medida.

Crítérios. É possível especificar o número máximo de iterações que o procedimento pode percorrer em seus cálculos. Também é possível selecionar um valor de critério de convergência. O algoritmo para a iteração se a diferença no ajuste total entre as duas últimas iterações for menor que o valor de convergência ou se o número máximo de iterações for atingido.

Gráficos por rótulo. Permite especificar se as variáveis e os rótulos de valor ou nomes e valores de variáveis serão usados nos gráficos. Também é possível especificar um comprimento máximo para rótulos.

Dimensões de gráfico. Permite controlar as dimensões exibidas na saída.

- **Exibir todas as dimensões na solução.** Todas as dimensões na solução são exibidas em uma matriz de gráfico de dispersão.
- **Restringir o número de dimensões.** As dimensões exibidas são restritas a pares plotados. Se você restringir as dimensões, deverá selecionar as dimensões mais baixas e mais altas a serem plotadas. A dimensão mais baixa pode ir de 1 ao número de dimensões na solução menos 1 e é plotada com relação às dimensões mais altas. O valor da dimensão mais alta pode ir de 2 ao número de dimensões na solução e indica a dimensão mais alta a ser usada na plotagem dos pares de dimensão. Esta especificação se aplica a todos os gráficos multidimensionais solicitados.

Configuração. É possível ler dados de um arquivo que contém as coordenadas de uma configuração. A primeira variável no arquivo contém as coordenadas para a primeira dimensão, a segunda variável deve conter as coordenadas para a segunda dimensão, e assim por diante.

- **Inicial.** A configuração no arquivo especificado será usada como o ponto de início da análise.
- **Fixo.** A configuração no arquivo especificado será usada para ajuste nas variáveis. As variáveis que são ajustadas devem ser selecionadas como variáveis de análise, mas, porque a configuração é fixa, elas são tratadas como variáveis complementares (de modo que não precisam ser selecionadas como variáveis complementares).

Saída de análise de múltipla correspondência

A caixa de diálogo de Saída permite produzir tabelas para escores do objeto, medidas de discriminação, histórico de iteração, correlações de variáveis originais e transformadas, quantificações de categoria para variáveis selecionadas e estatísticas descritivas para variáveis selecionadas.

Escores do objeto. Exibe os escores do objeto, incluindo a massa, inércia e contribuições, e tem as seguintes opções:

- **Incluir categorias de.** Exibe os indicadores de categoria das variáveis de análise selecionadas.
- **Rotular escores do objeto por.** A partir da lista de variáveis que são especificadas como variáveis de rotulagem, é possível selecionar uma para rotular os objetos.

Medidas de discriminação. Exibe as medidas de discriminação por variável e por dimensão.

Histórico de iteração. Para cada iteração, a variação contabilizada, a perda e o aumento na variação contabilizada são mostrados.

Correlações de variáveis originais. Mostra a matriz de correlações das variáveis originais e os autovalores desta matriz.

Correlações de variáveis transformadas. Mostra a matriz de correlações das variáveis transformadas (idealmente escalada) e os autovalores desta matriz.

Quantificações e contribuições de categoria. Dá as quantificações de categoria (coordenadas), incluindo massa, inércia e contribuições, para cada dimensão da(s) variável(is) selecionada(s).

Nota: as coordenadas e contribuições (incluindo a massa e a inércia) são exibidas em estratos separadas da saída da tabela dinâmica, com as coordenadas mostradas por padrão. Para exibir as contribuições, ative (clique duas vezes) a tabela e selecione Contribuições a partir da lista suspensa Camada.

Estatísticas descritivas. Exibe frequências, número de valores omissos e o modo da(s) variável(is) selecionada(s).

Salvamento da análise de múltipla correspondência

A caixa de diálogo Salvar permite salvar dados discretizados, escores do objeto e valores transformados em um arquivo de dados ou conjunto de dados externo do IBM SPSS Statistics na sessão atual. Também é possível salvar os valores transformados e escores do objeto no conjunto de dados ativo.

- Os conjuntos de dados estão disponíveis durante a sessão atual, mas não estão disponíveis nas sessões subsequentes, a menos que sejam salvos explicitamente como arquivos de dados. Os nomes dos conjuntos de dados devem aderir às regras de nomenclatura de variáveis.
- Os nomes dos arquivos ou os nomes do conjunto de dados devem ser diferentes para cada tipo de dado salvo.
- Se você salvar escores do objeto ou valores transformados no conjunto de dados ativo, será possível especificar o número de múltiplas dimensões nominais.

Gráficos de objetos de múltiplas análises de correspondência

A caixa de diálogo Gráficos de objeto permite especificar os tipos de gráficos que você deseja e as variáveis a serem plotadas.

Pontos de objeto. Um gráfico dos pontos de objeto é exibido.

Objetos e centroides (biplot). Os pontos de objetos são plotados com os centroides da variável.

Variáveis do biplot. É possível escolher usar todas as variáveis para os biplots ou selecionar um subconjunto.

Rotulas objetos. É possível escolher ter objetos rotulados com as categorias das variáveis selecionadas (deve-se escolher valores indicadores de categoria ou rótulos de valores na caixa de diálogo Opções) ou com seus números de caso. Um gráfico é produzido por variável se a **Variável** for selecionada.

Gráficos da variável de análise de múltiplas correspondências

A caixa de diálogo Gráficos da Variável permite especificar os tipos de gráficos que você deseja e as variáveis a serem plotadas.

Gráficos de categoria. Para cada variável selecionada, um gráfico das coordenadas do centroide é plotado. As categorias estão nos centroides dos objetos nas categorias particulares.

Juntar gráficos de categoria. Este é um gráfico único das coordenadas do centroide de cada variável selecionada.

Gráficos de transformação. Exibe um gráfico das qualificações da categoria ideal versus os indicadores de categoria. É possível especificar o número de dimensões; um gráfico será gerado para cada dimensão. Também é possível escolher exibir gráficos de resíduos para cada variável selecionada.

Medidas de discriminação. Produz um gráfico único das medidas de discriminação para as variáveis selecionadas.

Recursos adicionais do comando MULTIPLE CORRESPONDENCE

É possível customizar sua Análise de Múltipla Correspondência se você colar suas seleções em uma janela de sintaxe e editar a sintaxe de comando MULTIPLE CORRESPONDENCE resultante. O idioma da sintaxe de comando também permite:

- Especifique nomes de raiz para as variáveis transformadas, escores de objeto e aproximações ao salvá-las para o conjunto de dados ativo (com o subcomando SAVE).
- Especifique um comprimento máximo para rótulos para cada gráfico separadamente (com o subcomando PLOT).
- Especifique uma lista de variáveis separada para gráficos de resíduos (com o subcomando PLOT).

Consulte a *Referência da sintaxe de comando* para obter informações de sintaxe completa.

Ajuste de escala multidimensional (PROXSCAL)

O ajuste de escala multidimensional tenta localizar a estrutura em um conjunto de medidas de proximidade entre objetos. Este processo é executado ao designar observações a localizações específicas em um espaço conceitual de baixa dimensão, de modo que as distâncias entre pontos no espaço correspondam à determinadas (dis)similaridades tão próximas quanto possível. O resultado é uma representação de quadrados mínimos dos objetos nesse espaço de baixa dimensão, que, em muitos casos, vai ajudá-lo a entender melhor seus dados.

Exemplo. O ajuste de escala multidimensional pode ser muito útil em determinar relacionamentos perceptivos. Por exemplo, ao considerar sua imagem do produto, é possível conduzir uma pesquisa para obter um conjunto de dados que descreva a similaridade (ou proximidade) percebida de seu produto com aqueles de seus concorrentes. Usando essas proximidades e variáveis independentes (como preço), é possível tentar determinar quais variáveis são importantes para como as pessoas veem esses produtos, e é possível ajustar sua imagem de acordo com isso.

Estatísticas e gráficos. Histórico de iteração, medidas de estresse, decomposição de estresse, coordenadas do espaço comum, distâncias do objeto dentro da configuração final, ponderações de espaço individual, espaços individuais, proximidades transformadas, variáveis independentes transformadas, gráficos de estresse, gráficos de dispersão de espaço comum, gráficos de dispersão de ponderação de espaço individual, gráficos de dispersão de espaços individuais, gráficos de transformação, gráficos de resíduos de Shepard e gráficos de transformação de variáveis independentes.

Considerações de dados de ajuste de escala multidimensional

Dados. Os dados podem ser fornecidos na forma de matrizes de proximidade ou variáveis que são convertidas em matrizes de proximidade. As matrizes podem ser formatadas em linhas em linhas ou entre linhas. As proximidades podem ser tratadas nos níveis e ajuste de escala de razão, intervalo, ordinal ou spline.

Suposições. Pelo menos três variáveis devem ser especificadas. O número de dimensões não pode exceder o número de objetos menos um. A redução de dimensionalidade é omitida se combinada com vários inícios aleatórios. Se apenas uma origem for especificada, todos os modelos serão equivalentes ao modelo de identidade; portanto, a análise é padronizada para o modelo identidade.

Procedimentos relacionados. Ajustar a escala de todas as variáveis no nível numérico corresponde a análise de ajuste de escala multidimensional padrão.

Para obter um ajuste de escala multidimensional

1. Nos menus, escolha:

Analisar > Escala > Ajuste de escala multidimensional (PROXSCAL)...

Isto abre a caixa de diálogo Formato de dados.

2. Especifique o formato de seus dados:

Formato de Dados. Especifique se seus dados consistem de medidas de proximidades ou se deseja criar proximidades a partir dos dados.

Número de origens. Se seus dados são proximidades, especifique se você tem uma única origem ou várias origens de medidas de proximidade.

Uma origem. Se houver uma origem de proximidades, especifique se seu conjunto de dados é formatado com as proximidades em uma matriz entre as colunas ou em uma única coluna com duas variáveis separadas para identificar a linha e a coluna de cada proximidade.

- *As proximidades estão em uma matriz entre as colunas.* A matriz de proximidade é difundida para uma série de colunas igual ao número de objetos. Isso leva à caixa de diálogo Proximidades de Matrizes entre as Colunas.
- *As proximidades estão em uma única coluna.* A matriz de proximidade é reduzida em uma única coluna, ou variável. Duas variáveis adicionais, que identificam a linha e coluna para cada célula, são necessárias. Isso leva às Proximidades na caixa de diálogo Uma Coluna.

Várias origens. Se houver várias origens de proximidades, especifique se o conjunto de dados é formatado com as proximidades em matrizes empilhadas entre colunas, em várias colunas com uma origem por coluna ou em uma única coluna.

- *As proximidades estão em matrizes empilhadas entre as colunas.* As matrizes de proximidade são difundidas para uma série de colunas igual ao número de objetos e são empilhadas uma sobre a

outra através de um número de linhas igual ao número de objetos vezes o número de origens. Isso leva à caixa de diálogo Proximidades de Matrizes entre as Colunas.

- *As proximidades estão em colunas, uma origem por coluna.* As matrizes de proximidade são reduzidas em diversas colunas, ou variáveis. Duas variáveis adicionais, que identificam a linha e coluna para cada célula, são necessárias. Isso leva à caixa de diálogo Proximidades nas Colunas.
- *As proximidades são empilhadas em uma única coluna.* As matrizes de proximidade são reduzidas em uma única coluna, ou variável. Três variáveis adicionais, que identificam a linha, a coluna e a origem para cada célula, são necessárias. Isso leva às Proximidades na caixa de diálogo Uma Coluna.

3. Clique em **Definir**.

Proximidades em matrizes em colunas

Se você selecionar as proximidades no modelo de dados das matrizes para uma origem ou múltiplas origens na caixa de diálogo Formato de Dados, então, faça o seguinte:

1. Selecione três ou mais variáveis de proximidades. (Tenha certeza de que a ordem das variáveis na lista corresponde à ordem das colunas das proximidades.)
2. Opcionalmente, selecione um número de variáveis de ponderações igual ao número de variáveis de proximidades. (Tenha certeza de que a ordem das ponderações corresponde à ordem das proximidades que elas ponderam.)
3. Opcionalmente, se houver múltiplas origens, selecione uma variável de origens. (O número de casos em cada variável de proximidades deve ser igual ao número de variáveis de proximidades vezes o número de origens.)

Além disso, é possível definir um modelo para o ajuste de escala multidimensional, colocar restrições sobre o espaço comum, definir critérios de convergência, especificar a configuração inicial a ser usada e escolher gráficos e de saída.

Proximidades em colunas

Se você selecionar o modelo de múltiplas colunas para múltiplas origens na caixa de diálogo Formato de dados, faça o seguinte:

1. Selecionar duas ou mais variáveis de proximidade. (Cada variável é assumida para ser uma matriz de proximidades de uma origem separada.)
2. Selecione uma variável de linha para definir as localizações de linha para as proximidades em cada variável de proximidades.
3. Selecione uma variável de coluna para definir as localizações de coluna para as proximidades em cada variável de proximidades. (As células da matriz de proximidade que não recebem uma designação de linha/coluna são tratadas como omissas.)
4. Opcionalmente, selecione um número de variáveis de ponderações igual ao número de variáveis de proximidades.

Além disso, é possível definir um modelo para o ajuste de escala multidimensional, colocar restrições sobre o espaço comum, definir critérios de convergência, especificar a configuração inicial a ser usada e escolher gráficos e de saída.

Proximidades em uma coluna

Se você selecionar o modelo de uma coluna para uma origem ou várias origens na caixa de diálogo Formato de dados, faça o seguinte:

1. Selecione uma variável de proximidades. (é assumido como sendo uma ou mais matrizes de proximidades.)
2. Selecione uma variável de linha para definir os locais de linha para as proximidades na variável de proximidade.

3. Selecione uma variável da coluna para definir os locais da coluna para as proximidades na variável de proximidade.
4. Se houver várias origens, selecione uma variável de origens. (Para cada origem, as células da matriz de proximidade que não recebem uma designação de linha/coluna são tratadas como omissas.)
5. Opcionalmente, selecione uma variável de ponderação.

Além disso, é possível definir um modelo para o ajuste de escala multidimensional, colocar restrições sobre o espaço comum, definir critérios de convergência, especificar a configuração inicial a ser usada e escolher gráficos e de saída.

Criar proximidades a partir de dados

Se você escolher criar proximidades a partir de dados na caixa de diálogo Formato de dados, então, faça o seguinte:

1. Se você criar distâncias entre variáveis (consulte a caixa de diálogo Criar Medida a partir de Dados), selecione, pelo menos, três variáveis. Essas variáveis serão usadas para criar a matriz de proximidade (ou matrizes, se houver várias origens). Se você criar distâncias entre casos, apenas uma variável será necessária.
2. Se houver várias origens, selecione uma variável de origens.
3. Opcionalmente, escolha uma medida para criar proximidades.

Além disso, é possível definir um modelo para o ajuste de escala multidimensional, colocar restrições sobre o espaço comum, definir critérios de convergência, especificar a configuração inicial a ser usada e escolher gráficos e de saída.

Criar medida a partir de dados

O ajuste de escala multidimensional usa dados de dissimilaridade para criar uma solução de ajuste de escala. Se os seus dados são dados multivariados (valores das variáveis medidas), deve-se criar dados de dissimilaridade, a fim de calcular uma solução de ajuste de escala multidimensional. É possível especificar os detalhes da criação de medidas de dissimilaridade a partir de seus dados.

Medida. Permite especificar a medida de dissimilaridade para sua análise. Selecione uma alternativa a partir do grupo Medida correspondente ao seu tipo de dados e, em seguida, selecione uma das medidas a partir da lista suspensa correspondente ao tipo de medida. As alternativas disponíveis são:

- **Intervalo.** Distância euclidiana, distância euclidiana quadrada, Chebychev, bloco, Minkowski ou customizado.
- **Contagem.** Medida qui-quadrado ou medida fi-quadrado.
- **Binário.** Distância euclidiana, distância euclidiana quadrada, diferença de tamanho, diferença de padrão, variação ou Lance e Williams.

Criar matriz de distância. Permite escolher a unidade de análise. As alternativas são entre variáveis ou entre casos.

Transformar valores. Em determinados casos, como quando variáveis são medidas em escalas muito diferentes, deseja-se padronizar valores antes de calcular proximidades (não aplicável para dados binários). Selecione um método de padronização a partir da lista suspensa Padronizar (se nenhuma padronização for requerida, selecione **Nenhum**).

Definir um modelo de ajuste de escala multidimensional

A caixa de diálogo Modelo permite especificar um modelo de ajuste de escala, seu número máximo e mínimo de dimensões, a estrutura da matriz de proximidade, a transformação para usar nas proximidades e se as proximidades são transformadas dentro de cada origem separadamente ou incondicionalmente na origem.

Modelo de ajuste de escala. Escolha a partir das alternativas a seguir:

- **Identidade.** Todas as origens têm a mesma configuração.
- **Euclidiana ponderada.** Este modelo é um modelo de diferenças individuais. Cada origem tem um espaço individual no qual cada dimensão do espaço comum é ponderada de forma diferente.
- **Euclidiana generalizada.** Este modelo é um modelo de diferenças individuais. Cada origem tem um espaço individual que é igual a uma rotação do espaço comum, seguido por uma ponderação de diferencial das dimensões.
- **Ranqueamento reduzido.** Este modelo é um modelo euclidiano generalizado para o qual é possível especificar o ranqueamento do espaço individual. Deve-se especificar um ranqueamento que é maior ou igual a 1 e menor que o número máximo de dimensões.

Forma. Especifique se as proximidades devem ser levadas a partir da parte triangular inferior ou da parte triangular superior da matriz de proximidade. É possível especificar que a matriz integral a ser usada, no caso de a soma ponderada da parte triangular superior e na parte triangular inferior ser analisada. Em qualquer caso, a matriz completa deve ser especificada, incluindo diagonal, embora apenas as partes especificadas sejam utilizadas.

Proximidades. Especifique se sua matriz de proximidade contém medidas de similaridade ou dissimilaridade.

Transformações de proximidade. Escolha a partir das alternativas a seguir:

- **Razão.** As proximidades transformadas são proporcionais às proximidades originais. Isso é permitido apenas para proximidades com valor positivo.
- **Intervalo.** As proximidades transformadas são proporcionais às proximidades originais, mais um termo de interceptação. O intercepto assegura que todas as proximidades sejam positivas.
- **Ordinal.** As proximidades transformadas têm a mesma ordem que as proximidades originais. Especifique se as proximidades empatadas devem ser mantidas empatadas ou se é permitido que elas se tornem desempatadas.
- **Spline.** As proximidades transformadas são uma transformação polinomial suave em trechos não decrescentes. Especifique o grau do polinomial e o número de nós internos.

Aplicar transformações. Especifique se apenas as proximidades dentro de cada origem são comparadas umas com as outras ou se as comparações são incondicionais na origem.

Dimensões. Por padrão, uma solução é calculada em duas dimensões (Mínimo = 2, Máximo = 2). Escolha um número inteiro mínimo e máximo de 1 ao número de objetos menos 1 (desde que o mínimo seja menor ou igual ao máximo). O procedimento calcula uma solução no máximo de dimensões e, em seguida, reduz a dimensionalidade em etapas até o mínimo ser atingido.

Restrições de ajuste de escala multidimensional

A caixa de diálogo Restrições permite colocar restrições no espaço comum.

Restrições no espaço comum. Especifique o tipo de restrição desejado.

- **Sem restrições.** Nenhuma restrição é colocada no espaço comum.
- **Algumas coordenadas fixas.** A primeira variável selecionada contém as coordenadas dos objetos na primeira dimensão, a segunda variável corresponde às coordenadas na segunda dimensão, e assim por diante. Um valor omissivo indica que uma coordenada em uma dimensão é livre. O número de variáveis selecionadas deve ser igual ao número máximo de dimensões solicitadas.
- **Combinação linear de variáveis independentes.** O espaço comum é restrito a ser uma combinação linear das variáveis selecionadas.

Variáveis de restrição. Selecione as variáveis que definem as restrições no espaço comum. Se tiver especificado uma combinação linear, especifique uma transformação de intervalo, nominal, ordinal ou de spline para as variáveis de restrição. Em ambos os casos, o número de casos para cada variável deve ser igual ao número de objetos.

Opções de ajuste de escala multidimensional

A caixa de diálogo Opções permite selecionar o estilo de configuração inicial, especificar critérios de iteração e convergência e selecionar atualizações padrão ou afrouxadas.

Configuração inicial. Escolha uma das alternativas a seguir:

- **Simplex.** Os objetos estão na mesma distância um do outro na dimensão máxima. Uma iteração é executada para melhorar esta configuração de alta dimensão, seguida por uma operação de redução de dimensão para obter uma configuração inicial que tenha o número máximo de dimensões especificadas na caixa de diálogo Modelo.
- **Torgerson.** Uma solução de ajuste de escala clássica é usada como a configuração inicial.
- **Início aleatório único.** Uma configuração é escolhida aleatoriamente.
- **Múltiplos inícios aleatórios.** Várias configurações são escolhidas aleatoriamente, e a configuração com estresse bruto normalizado mais baixo é usada como a configuração inicial.
- **Customizado.** Selecione variáveis que contém as coordenadas de sua própria configuração inicial. O número de variáveis selecionadas deve ser igual ao número máximo de dimensões especificadas, com a primeira variável correspondendo às coordenadas na dimensão 1, a segunda variável correspondendo às coordenadas na dimensão 2, e assim por diante. O número de casos em cada variável deve ser igual ao número de objetos.

CrITÉRIOS de iteraÇÃO. Especifique os valores de critérios de iteraÇão.

- **Convergência de estresse.** O algoritmo parará a iteraÇão quando a diferença em valores de estresse penalizado normalizado consecutivos for menor que o número especificado aqui, que deve estar entre 0.0 e 1.0.
- **Estresse mÍNIMO.** O algoritmo parará quando o estresse bruto normalizado cair abaixo do número especificado aqui, que deve estar entre 0.0 e 1.0.
- **Máximo de iteraÇões.** O algoritmo executará o número de iteraÇões especificadas, a menos que um dos critérios acima seja satisfeito primeiro.
- **Usar atualizações afrouxadas.** As atualizações afrouxadas acelerarão o algoritmo; estas atualizações não podem ser usadas com modelos diferentes do modelo de identidade ou usadas com restrições.

Gráficos de ajuste de escala multidimensional, versão 1

A caixa de diálogo Gráficos permite especificar quais gráficos serão produzidos. Este tópico descreve a caixa de diálogo Gráficos se você tiver o formato de dados Proximidades em Colunas. Para os gráficos **Ponderações de espaço individual**, **Original vs. proximidades transformadas** e **Proximidades transformadas versus distâncias**, especifique as origens para as quais os gráficos devem ser produzidos. A lista de origens disponíveis é a lista de variáveis de proximidades na caixa de diálogo principal.

Estresse. Um gráfico é produzido do estresse bruto normalizado versus dimensões. Este gráfico é produzido apenas se o número máximo de dimensões for maior que o número mínimo de dimensões.

Espaço comum. Uma matriz de gráfico de dispersão das coordenadas do espaço comum é exibida.

Espaços individuais. Para cada origem, as coordenadas dos espaços individuais são exibidas nas matrizes dos gráficos de dispersão. Isso só é possível se um dos modelos de diferenças individuais for especificado na caixa de diálogo Modelo.

Ponderações de espaço individual. Um gráfico de dispersão é produzido das ponderações de espaços individuais. Isso só é possível se um dos modelos de diferenças individuais for especificado na caixa de

diálogo Modelo. Para o modelo euclidiano ponderado, as ponderações são impressas em gráficos, com uma dimensão em cada eixo. Para o modelo Euclidiano generalizado, um gráfico é produzido por dimensão, indicando tanto a rotação como o peso desta dimensão. O modelo de ranqueamento reduzido produz o mesmo gráfico que o modelo Euclidiano generalizado, mas reduz o número de dimensões para os espaços individuais.

Proximidades originais vs. transformadas. Os gráficos são produzidos das proximidades originais versus as proximidades transformadas.

Proximidades transformadas versus distâncias. As proximidades transformadas versus as distâncias são plotadas.

Variáveis independentes transformadas. Gráficos de transformação são produzidos para variáveis independentes.

Correlações de variável e dimensão. Um gráfico de correlações entre as variáveis independentes e as dimensões do espaço comum é exibido.

Gráficos de ajuste de escala multidimensional, versão 2

A caixa de diálogo Gráficos permite especificar quais gráficos serão produzidos. Este tópico descreve a caixa de diálogo Gráficos se seu formato de dados for algo diferente de Proximidades em Colunas. Para os gráficos **Ponderações de espaço individual**, **Original vs. proximidades transformadas** e **Proximidades transformadas versus distâncias**, especifique as origens para as quais os gráficos devem ser produzidos. Os números de origens inseridos devem ser valores da variável de origens especificada na caixa de diálogo principal e deve variar de 1 ao número de origens.

Saída de ajuste de escala multidimensional

A caixa de diálogo Saída permite controlar a quantidade de saídas exibidas e salvar algumas em arquivos separados.

Exibição. Selecione um ou mais dos itens a seguir para exibição:

- **Coordenadas de espaço comum.** Exibe as coordenadas do espaço comum.
- **Coordenadas de espaço individual.** As coordenadas dos espaços individuais são exibidas apenas se o modelo não for o modelo identidade.
- **Ponderações de espaço individual.** Exibe as ponderações de espaço individual apenas se um dos modelos de diferenças individuais for especificado. Dependendo do modelo, as ponderações de espaço são decompostas em ponderações de rotação e ponderações de dimensão, que também são exibidas.
- **Distâncias.** Exibe as distâncias entre os objetos na configuração.
- **Proximidades transformadas.** Exibe as proximidades transformadas entre os objetos na configuração.
- **Dados de entrada.** Inclui as proximidades originais e, se presente, as ponderações de dados, a configuração inicial e as coordenadas fixas das variáveis independentes.
- **Estresse para inícios aleatórios.** Exibe o valor de valor semente de número aleatório e de estresse bruto normalizado de cada início aleatório.
- **Histórico de iteração.** Exibe o histórico de iterações do algoritmo principal.
- **Múltiplas medidas de estresse.** Exibe os diferentes valores de estresse. A tabela contém valores para estresse bruto normalizado, Stress-I, Stress-II, S-Stress, Dispersion Accounted For (DAF) e Coeficiente de Congruência de Tucker.
- **Decomposição de estresse.** Exibe uma decomposição de objetos e origens do estresse bruto normalizado final, incluindo a média por objeto e a média por origem.
- **Variáveis independentes transformadas.** Se uma restrição de combinação linear foi selecionada, as variáveis independentes transformadas e as ponderações de regressão correspondentes serão exibidas.

- **Correlações de variável e dimensão.** Se uma restrição de combinação linear foi selecionada, as correlações entre as variáveis independentes e as dimensões do espaço comum serão exibidas.

Salvar para novo arquivo. É possível salvar as coordenadas do espaço comum, ponderações do espaço individual, distâncias, proximidades transformadas e variáveis independentes transformadas em arquivos de dados do IBM SPSS Statistics separados.

Recursos adicionais de comandos do PROXSCAL

É possível customizar seu ajuste de escala multidimensional de análise de proximidades se você colar suas seleções em uma janela de sintaxe e editar a sintaxe de comando PROXSCAL resultante. O idioma da sintaxe de comando também permite:

- Especifique listas de variáveis separadas para gráficos de transformações e resíduos (com o subcomando PLOT).
- Especifique listas de origens separadas para ponderações de espaços individuais, transformações e gráficos de resíduos (com o subcomando PLOT).
- Especifique um subconjunto dos gráficos de transformação de variáveis independentes a serem exibidos (com o subcomando PLOT).

Consulte a *Referência da sintaxe de comando* para obter informações de sintaxe completa.

Desdobramento multidimensional (PREFSCAL)

O procedimento Desdobramento Multidimensional tenta localizar uma escala quantitativa comum que permite examinar visualmente os relacionamentos entre dois conjuntos de objetos.

Exemplos. Você pediu a 21 indivíduos que classificassem 15 itens de café da manhã na ordem de preferência, de 1 a 15. Usando o Desdobramento Multidimensional, é possível determinar que os indivíduos discriminam entre itens de café da manhã de duas formas primárias: entre pães macios e duros e entre itens que engorda e que não engordam.

Alternativamente, você solicitou a um grupo de motoristas que classificassem 26 modelos de carros em 10 atributos em uma escala de 6 pontos que varia de 1 = "não é tudo verdade" a 6 = "muito verdadeiro". Obtidas as médias sobre os indivíduos, os valores são tomados como similaridades. Usando o Desdobramento Multidimensional, você encontra armazenamentos em cluster de modelos similares e os atributos com os quais eles são mais proximamente associados.

Estatísticas e gráficos. O procedimento Desdobramento Multidimensional pode produzir um histórico de iteração, medidas de estresse, decomposição de estresse, coordenadas do espaço comum, distâncias de objeto dentro da configuração final, ponderações de espaços individuais, espaços individuais, proximidades transformadas, plotagens de estresse, gráficos de dispersão de espaço comum, gráficos de dispersão de ponderação de espaço individual, gráficos de dispersão de espaços individuais, gráficos de transformação e gráficos residuais de Shepard.

Considerações de dados de desdobramento multidimensional

Dados. Os dados são fornecidos na forma de matrizes de proximidade retangular. Cada coluna é considerado um objeto de coluna separado. Cada linha de uma matriz de proximidade é considerada um objeto de linha separado. Quando há várias origens de proximidades, as matrizes são empilhadas.

Suposições. Pelo menos duas variáveis devem ser especificadas. O número de dimensões na solução não pode exceder o número de objetos menos um. Se apenas uma origem for especificada, todos os modelos serão equivalentes ao modelo de identidade; portanto, a análise é padronizada para o modelo identidade.

Para obter um Desdobramento multidimensional

1. Nos menus, escolha:
Analisar > Escala > Desdobramento Multidimensional (PREFSCAL)...
2. Selecione duas ou mais variáveis que identificam as colunas na matriz de proximidade retangular. Cada variável representa um objeto de coluna separado.
3. Opcionalmente, selecione um número de variáveis de ponderações igual ao objeto de variáveis do objeto de coluna. A ordem das variáveis de ponderações deve corresponder à ordem dos objetos da coluna que eles ponderam.
4. Opcionalmente, selecione uma variável de linhas. Os valores (ou rótulos de valor) desta variável são usados para rotular objetos de linha na saída.
5. Se houver origens múltiplas, selecione, opcionalmente, uma variável de origem. O número de casos no arquivo de dados deve ser igual ao número de objetos de linha vezes o número de origens.

Além disso, é possível definir um modelo para o desdobramento multidimensional, colocar restrições sobre o espaço comum, definir critérios de convergência, especificar a configuração inicial a ser usada e escolher gráficos e saída.

Definir um modelo de Desdobramento multidimensional

A caixa de diálogo Modelo permite especificar um modelo de ajuste de escala, seu número máximo e mínimo de dimensões, a estrutura da matriz de proximidade, a transformação para usar nas proximidades e se as proximidades são transformadas condicionalmente sobre a linha, condicionalmente sobre a origem ou incondicionalmente sobre a origem.

Modelo de ajuste de escala. Escolha a partir das alternativas a seguir:

- **Identidade.** Todas as origens têm a mesma configuração.
- **Euclidiana ponderada.** Este modelo é um modelo de diferenças individuais. Cada origem tem um espaço individual no qual cada dimensão do espaço comum é ponderada de forma diferente.
- **Euclidiana generalizada.** Este modelo é um modelo de diferenças individuais. Cada origem tem um espaço individual que é igual a uma rotação do espaço comum, seguido por uma ponderação de diferencial das dimensões.

Proximidades. Especifique se sua matriz de proximidade contém medidas de similaridade ou dissimilaridade.

Dimensões. Por padrão, uma solução é calculada em duas dimensões (Mínimo = 2, Máximo = 2). É possível escolher um número inteiro mínimo e máximo de 1 ao número de objetos menos 1, desde que o mínimo seja menor ou igual ao máximo. O procedimento calcula uma solução na dimensionalidade máxima e, em seguida, reduz a dimensionalidade em etapas até o mínimo ser atingido.

Transformações de proximidade. Escolha a partir das alternativas a seguir:

- **Nenhum.** As proximidades não são transformadas. É possível selecionar opcionalmente **Incluir intercepto**, em casos em que as proximidades podem ser deslocadas por um termo constante.
- **Linear.** As proximidades transformadas são proporcionais às proximidades originais; isto é, a função de transformação estima uma inclinação e o intercepto é fixado em 0. Isto é também chamado de uma transformação de razão. É possível selecionar opcionalmente **Incluir intercepto**, em casos em que as proximidades também podem ser deslocadas por um termo constante. Isto também é chamado de uma transformação de intervalo.
- **Spline.** As proximidades transformadas são uma transformação polinomial suave em trechos não decrescentes. É possível especificar o grau do polinomial e o número de nós internos. É possível selecionar opcionalmente **Incluir intercepto**, em casos em que as proximidades também podem ser deslocadas por um termo constante.

- **Suave.** As proximidades transformadas têm a mesma ordem que as proximidades originais, incluindo uma restrição que leva as diferenças entre valores subsequentes em consideração. O resultado é uma transformação "ordinal suave". É possível especificar se proximidades empatadas devem ser mantidas empatadas ou se é permitido desempatá-las.
- **Ordinal.** As proximidades transformadas têm a mesma ordem que as proximidades originais. É possível especificar se proximidades empatadas devem ser mantidas empatadas ou se é permitido desempatá-las.

Aplicar transformações. Especifique se apenas as proximidades dentro de cada linha são comparadas umas com as outras ou se apenas proximidades dentro de cada origem são comparadas umas com as outras, ou se as comparações são incondicionais na linha ou origem; isto é, se as transformações são executadas por linha, por origem ou por todas as proximidades de uma vez.

Restrições de desdobramento multidimensional

A caixa de diálogo Restrições permite colocar restrições no espaço comum.

Restrições no espaço comum. É possível escolher corrigir as coordenadas de objetos de linha e/ou coluna no espaço comum.

Variáveis de restrição de linha/coluna. Escolha o arquivo que contém as restrições e selecione as variáveis que definem as restrições no espaço comum. A primeira variável selecionada contém as coordenadas dos objetos na primeira dimensão, a segunda variável corresponde às coordenadas na segunda dimensão, e assim por diante. Um valor omisso indica que uma coordenada em uma dimensão é livre. O número de variáveis selecionadas deve ser igual ao número máximo de dimensões solicitadas. O número de casos para cada variável deve ser igual ao número de objetos.

Opções de desdobramento multidimensional

A caixa de diálogo Opções permite selecionar o estilo de configuração inicial, especificar critérios de iteração e convergência e configurar o termo de penalidade para estresse.

Configuração inicial. Escolha uma das alternativas a seguir:

- **Clássico.** A matriz de proximidade retangular é usada para complementar os intrabloques (valores entre linhas e entre colunas) da matriz MDS simétrica completa. Uma vez que a matriz completa é formada, uma solução de ajuste de escala clássica é usada como a configuração inicial. Os intrabloques podem ser preenchidos por meio de entrada, usando a desigualdade do triângulo ou as distâncias de Spearman.
- **Ross-Cliff.** O início de Ross-Cliff usa os resultados de uma decomposição em valores singulares na matriz de proximidade centrada dupla e quadrada como os valores iniciais para os objetos de linha e coluna.
- **Correspondência.** O início da correspondência usa os resultados de uma análise de correspondência sobre os dados reversos (similaridades em vez de dissimilaridades), com normalização simétrica de escores de linha e coluna.
- **Centroides.** O procedimento inicia posicionando os objetos de linha na configuração, usando uma decomposição de autovalor. Então, os objetos de coluna são posicionados no centroide das opções especificadas. Para o número de opções, especifique um número inteiro positivo entre 1 e o número de variáveis de proximidades.
- **Múltiplos inícios aleatórios.** As soluções são calculadas para várias configurações iniciais escolhidas aleatoriamente e aquela com estresse penalizado mais baixo é mostrada como a melhor solução.
- **Customizado.** É possível selecionar variáveis que contém as coordenadas de sua própria configuração inicial. O número de variáveis selecionadas deve ser igual ao número máximo de dimensões especificadas, com a primeira variável correspondendo às coordenadas na dimensão 1, a segunda variável correspondendo às coordenadas na dimensão 2, e assim por diante. O número de casos em cada variável deve ser igual ao número combinado de objetos de linha e coluna. As coordenadas de linha e coluna devem ser empilhadas, com as coordenadas de coluna após as coordenadas de linha.

Critérios de iteração. Especifique os valores de critérios de iteração.

- **Convergência de estresse.** O algoritmo parará a iteração quando a diferença relativa em valores de estresse penalizado consecutivos for menor que o número especificado aqui, que deve ser não negativo.
- **Estresse mínimo.** O algoritmo parará quando o estresse penalizado cair abaixo do número especificado aqui, que deve ser não negativo.
- **Máximo de iterações.** O algoritmo executará o número de iterações especificadas aqui, a menos que um dos critérios acima seja satisfeito primeiro.

Termo de penalidade. O algoritmo tenta minimizar o estresse penalizado, uma medida de qualidade de ajuste igual ao produto de Kruskal's Stress-I e um termo de penalidade baseado no coeficiente de variação das proximidades transformadas. Esses controles permitem configurar a intensidade e o intervalo do termo de penalidade.

- **Intensidade.** Quanto menor o valor do parâmetro de intensidade, maior a penalidade. Especifique um valor entre 0.0 e 1.0.
- **Intervalo.** Este parâmetro configura o momento em que a penalidade se torna ativa. Se configurada para 0.0, a penalidade é inativa. Aumentar o valor faz com que o algoritmo procure uma solução com variação maior entre as proximidades transformadas. Especifique um valor não negativo.

Gráficos de desdobramento multidimensional

A caixa de diálogo Gráficos permite especificar quais gráficos serão produzidos.

Gráficos. Os gráficos a seguir estão disponíveis:

- **Inícios múltiplos.** Exibe um histograma empilhado de estresse penalizado que exibe estresse e penalidade.
- **Espaço comum inicial.** Exibe uma matriz de gráfico de dispersão das coordenadas do espaço comum inicial.
- **Stress por dimensão.** Produz um gráfico de linha de estresse penalizado versus dimensionalidade. Este gráfico é produzido apenas se o número máximo de dimensões for maior que o número mínimo de dimensões.
- **Espaço comum final.** Uma matriz de gráfico de dispersão das coordenadas do espaço comum é exibida.
- **Ponderações de espaço.** Um gráfico de dispersão é produzido das ponderações de espaços individuais. Isso só é possível se um dos modelos de diferenças individuais for especificado na caixa de diálogo Modelo. Para o modelo euclidiano ponderado, os pesos para todas as fontes são exibidas em um gráfico, com uma dimensão em cada eixo. Para o modelo euclidiano generalizado, um gráfico é produzido por dimensão, indicando tanto a rotação como o peso desta dimensão para cada origem.
- **Espaços individuais.** Uma matriz de gráfico de dispersão das coordenadas do espaço individual de cada origem é exibida. Isso só é possível se um dos modelos de diferenças individuais for especificado na caixa de diálogo Modelo.
- **Gráficos de transformação.** Um gráfico de dispersão é produzido das proximidades originais versus das proximidades transformadas. Dependendo de como as transformações são aplicadas, uma cor separada é designada para cada linha ou origem. Uma transformação incondicional produz uma cor única.
- **Gráficos Shepard.** As proximidades originais versus as proximidades e distâncias transformadas. As distâncias são indicadas pelos pontos, e as proximidades transformadas são indicadas por uma linha. Dependendo de como as transformações são aplicadas, uma borda separada é produzida para cada linha ou origem. Uma transformação incondicional produz uma única borda.
- **Gráfico de dispersão de ajuste.** Um gráfico de dispersão das proximidades transformadas versus as distâncias é exibido. Uma cor separada é designada para cada origem se várias origens forem especificadas.

- **Gráficos de resíduos.** Um gráfico de dispersão das proximidades transformadas versus os resíduos (proximidades transformadas menos distâncias) é exibido. Uma cor separada é designada para cada origem se várias origens forem especificadas.

Estilos de objeto de linha. Estes dão ao usuário ainda mais controle da exibição de objetos de linha em gráficos. Os valores da variável de cores opcionais são usados para percorrer todas as cores. Os valores da variável de Favoritos opcionais são usados para percorrer todos os Favoritos possíveis.

Gráficos de origem. Para **Espaços individuais**, **Gráfico de dispersão de ajuste** e **Gráficos de resíduos**—e se as transformações forem aplicadas por origem, para **Gráficos de transformação** e **Gráficos Shepard**—é possível especificar as origens para as quais os gráficos devem ser produzidos. Os números de origens inseridos devem ser valores da variável de origens especificada na caixa de diálogo principal e ir de 1 ao número de origens.

Gráficos de linha. Se as transformações forem aplicadas por linhas, para **Gráficos de transformação** e **Gráficos Shepard**, será possível especificar a linha para a qual os gráficos devem ser produzidos. Os números de linhas inseridos devem ir de 1 ao número de linhas.

Saída de Desdobramento multidimensional

A caixa de diálogo Saída permite controlar a quantia de saídas exibidas e salvar algumas em arquivos separados.

Exibição. Selecione uma ou mais das seguintes para exibir:

- **Dados de entrada.** Inclui as proximidades originais e, se presente, as ponderações de dados, a configuração inicial e as coordenadas fixas.
- **Inícios múltiplos.** Exibe o valor de valor semente de número aleatório e de estresse penalizado de cada início aleatório.
- **Dados iniciais.** Exibe as coordenadas do espaço comum inicial.
- **Histórico de iteração.** Exibe o histórico de iterações do algoritmo principal.
- **Medidas de ajuste.** Exibe diferentes medidas. A tabela contém várias medidas de qualidades do ajuste, deficiência do ajuste, correlação, variação e não degeneração.
- **Decomposição de estresse.** Exibe uma decomposição de objetos, linhas e origens de estresse penalizado, incluindo médias de linha, coluna e origens e desvios padrão.
- **Proximidades transformadas.** Exibe as proximidades transformadas.
- **Espaço comum final.** Exibe as coordenadas do espaço comum.
- **Ponderações de espaço.** Exibe as ponderações de espaço individual. Esta opção está disponível apenas se um dos modelos de diferença individuais for especificado. Dependendo do modelo, as ponderações de espaço são decompostas em ponderações de rotação e ponderações de dimensão, que também são exibidas.
- **Espaços individuais.** As coordenadas dos espaços individuais são exibidas. Esta opção está disponível apenas se um dos modelos de diferença individuais for especificado.
- **Distâncias ajustadas.** Exibe as distâncias entre os objetos na configuração.

Salvar para novo arquivo. É possível salvar as coordenadas de espaço comum, ponderações de espaço individual, distâncias e proximidades transformadas em arquivos de dados separados do IBM SPSS Statistics.

Recursos adicionais do comando PREFSCAL

É possível customizar seu Desdobramento multidimensional de análises de proximidades se você colar suas seleções em uma janela de sintaxe e editar a sintaxe de comando PREFSCAL resultante. O idioma da sintaxe de comando também permite:

- Especifique várias listas de origem para espaços individuais, gráficos de dispersão de ajuste e gráficos de resíduos—e no caso de transformações condicionais da matriz, para gráficos de transformação e gráficos de Shepard—quando várias origens estão disponíveis (com o subcomando PLOT).
- Especifique várias listas de linha para gráficos de transformação e gráficos de Shepard no caso de transformações condicionais de linha (com o subcomando PLOT).
- Especifique um número de linhas, em vez de uma variável de ID de linha (com o subcomando INPUT).
- Especifique um número de origens, em vez de uma variável de ID de origem (com o subcomando INPUT).

Consulte a *Referência da sintaxe de comando* para obter informações de sintaxe completa.

Avisos

Essas informações foram desenvolvidas para produtos e serviços oferecidos nos Estados Unidos. Esse material pode estar disponível a partir da IBM em outros idiomas. No entanto, pode ser necessário possuir uma cópia do produto ou da versão do produto nesse idioma para acessá-lo.

É possível que a IBM não ofereça produtos, serviços ou recursos discutidos neste documento em outros países. Consulte um representante IBM local para obter informações sobre produtos e serviços disponíveis atualmente em sua área. Qualquer referência a produtos, programas ou serviços IBM não significa que apenas produtos, programas ou serviços IBM possam ser utilizados. Qualquer produto, programa ou serviço funcionalmente equivalente, que não infrinja nenhum direito de propriedade intelectual da IBM poderá ser utilizado em substituição a este produto, programa ou serviço. Entretanto, a avaliação e verificação da operação de qualquer produto, programa ou serviço não IBM são de responsabilidade do Cliente.

A IBM pode ter patentes ou solicitações de patentes pendentes relativas a assuntos tratados nesta publicação. O fornecimento desta publicação não lhe garante direito algum sobre tais patentes. Pedidos de licença podem ser enviados, por escrito, para:

Gerência de Relações Comerciais e Industriais da IBM Brasil
Av. Pasteur, 138-146
CEP 22290-240
Rio de Janeiro, RJ
Brasil

Para pedidos de licença relacionados a informações de DBCS (Conjunto de Caracteres de Byte Duplo), entre em contato com o Departamento de Propriedade Intelectual da IBM em seu país ou envie pedidos de licença, por escrito, para:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO "NO ESTADO EM QUE SE ENCONTRA", SEM GARANTIA DE NENHUM TIPO, SEJA EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS NÃO SE LIMITANDO ÀS GARANTIAS IMPLÍCITAS DE NÃO-VIOLAÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO PROPÓSITO. Alguns países não permitem a exclusão de garantias explícitas ou implícitas em certas transações; portanto, esta instrução pode não se aplicar ao Cliente.

Essas informações podem conter imprecisões técnicas ou erros tipográficos. São feitas alterações periódicas nas informações aqui contidas; tais alterações serão incorporadas em futuras edições desta publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar o(s) produto(s) e/ou programa(s) descritos nesta publicação, sem aviso prévio.

Qualquer referência nestas informações a websites não IBM são fornecidas apenas por conveniência e não representam de forma alguma um endosso a esses websites. Os materiais contidos nesses websites não fazem parte dos materiais para esse produto IBM e o uso desses websites é de inteira responsabilidade do Cliente.

A IBM por usar ou distribuir as informações fornecidas da forma que julgar apropriada sem incorrer em qualquer obrigação para com o Cliente.

Licenciados deste programa que desejam obter informações sobre o mesmo com o objetivo de permitir: (i) a troca de informações entre programas criados independentemente e outros programas (incluindo este) e (ii) o uso mútuo de informações trocadas, devem entrar em contato com:

Gerência de Relações Comerciais e Industriais da IBM Brasil
Av. Pasteur, 138-146
CEP 22290-240
Rio de Janeiro, RJ
Brasil

Tais informações podem estar disponíveis, sujeitas a termos e condições apropriadas, incluindo em alguns casos o pagamento de uma taxa.

O programa licenciado descrito nesta publicação e todo o material licenciado disponível são fornecidos pela IBM sob os termos do Contrato com o Cliente IBM, do Contrato Internacional de Licença do Programa IBM ou de qualquer outro contrato equivalente.

Os exemplos de dados de desempenho e do Cliente citados são apresentados apenas para propósitos ilustrativos. Resultados de desempenho reais podem variar dependendo das configurações específicas e das condições operacionais.

Informações relativas a produtos não IBM foram obtidas junto aos fornecedores dos respectivos produtos, de seus anúncios publicados ou de outras fontes disponíveis publicamente. A IBM não testou esses produtos e não pode confirmar a precisão de desempenho, compatibilidade nem qualquer outra reivindicação relacionada a produtos não IBM. Perguntas sobre os recursos de produtos não IBM devem ser endereçadas aos fornecedores desses produtos.

Instruções relativas à direção futura ou intento da IBM estão sujeitas a mudança ou retirada sem aviso e representam metas e objetivos apenas.

Estas informações contêm exemplos de dados e relatórios utilizados nas operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos podem incluir nomes de assuntos, empresas, marcas e produtos. Todos esses nomes são fictícios e qualquer semelhança com pessoas ou empresas reais é mera coincidência.

LICENÇA DE COPYRIGHT:

Estas informações contêm programas de aplicativos de amostra na linguagem fonte, ilustrando as técnicas de programação em diversas plataformas operacionais. O Cliente pode copiar, modificar e distribuir estes programas de amostra sem a necessidade de pagar à IBM, com objetivos de desenvolvimento, utilização, marketing ou distribuição de programas aplicativos em conformidade com a interface de programação de aplicativo para a plataforma operacional para a qual os programas de amostra são criados. Esses exemplos não foram testados completamente em todas as condições. Portanto, a IBM não pode garantir ou implicar a confiabilidade, manutenção ou função destes programas. Os programas de amostra são fornecidos "NO ESTADO EM QUE SE ENCONTRAM", sem garantia de qualquer tipo. A IBM não será responsabilizada por quaisquer danos decorrentes do uso dos programas de amostra.

Cada cópia ou parte destes programas de amostra ou qualquer trabalho derivado deve incluir um aviso de copyright com os dizeres:

© nome de sua empresa) (ano). Partes deste código são derivadas dos Programas de Amostra da IBM Corp.

Marcas comerciais

IBM, o logotipo IBM e ibm.com são marcas comerciais ou marcas registradas da International Business Machines Corp., registradas em muitos países no mundo todo. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. A lista atual de marcas comerciais da IBM está disponível na web em "Copyright and trademark information" em www.ibm.com/legal/copytrade.shtml.

Adobe, o logotipo Adobe, PostScript e o logotipo PostScript são marcas registradas ou marcas comerciais da Adobe Systems Incorporated nos Estados Unidos e/ou em outros países.

Intel, o logotipo Intel, Intel Inside, o logotipo Intel Inside, Intel Centrino, o logotipo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium são marcas comerciais ou marcas registradas da Intel Corporation ou de suas subsidiárias nos Estados Unidos e em outros países.

Linux é uma marca registrada da Linus Torvalds nos Estados Unidos, e/ou em outros países.

Microsoft, Windows, Windows NT e o logotipo Windows são marcas comerciais da Microsoft Corporation nos Estados Unidos e/ou em outros países.

UNIX é uma marca registrada da The Open Group nos Estados Unidos e em outros países.

Java e todas as marcas comerciais e logotipos baseados em Java são marcas comerciais ou marcas registradas da Oracle e/ou suas afiliadas.

Índice Remissivo

A

- ajuste
 - na Análise de correlação canônica não linear 25
- Análise de componentes principais para dados categóricos 16, 19
 - autoinicialização 22
 - nível de ajuste de escala ideal 17
 - recursos adicionais do comando 23
 - salvar variáveis 21
- Análise de correlação canônica não linear 23, 25
 - estatísticas 23
 - gráficos 23
 - recursos adicionais do comando 26
- Análise de correspondências 26, 28, 29, 30
 - estatísticas 26
 - gráficos 26
 - recursos adicionais do comando 30
- Análise de correspondências múltipla 31, 33
 - nível de ajuste de escala ideal 32
 - recursos adicionais do comando 35
 - salvar variáveis 34
- ANOVA
 - na Regressão Categórica 14
- atualizações afrouxadas
 - no Ajuste de escala multidimensional 40
- autoinicialização
 - Análise de componentes principais para dados categóricos 22

B

- biplots
 - em análise de correspondência 30
 - em análise de múltiplas correspondências 35
 - na análise de componentes principais para dados categóricos 21

C

- carregamentos de componente
 - na análise de componentes principais para dados categóricos 20
 - na Análise de correlação canônica não linear 25
- centroídes
 - na Análise de correlação canônica não linear 25
- coeficientes de regressão
 - na Regressão Categórica 14
- configuração inicial
 - na Análise de correlação canônica não linear 25
 - na Regressão Categórica 13

- configuração inicial (*continuação*)
 - no Ajuste de escala multidimensional 40
 - no desdobramento multidimensional 44
- coordenadas de espaço comum
 - no Ajuste de escala multidimensional 41
 - no desdobramento multidimensional 46
- coordenadas de espaço individual
 - no desdobramento multidimensional 46
- correlações
 - no Ajuste de escala multidimensional 41
- critérios de iteração
 - no Ajuste de escala multidimensional 40
 - no desdobramento multidimensional 44

D

- Desdobramento multidimensional 42
 - estatísticas 42
 - gráficos 42, 45
 - model 43
 - opções 44
 - recursos adicionais do comando 46
 - restrições em espaço comum 44
 - saída 46
- Diagramas Shepard
 - no desdobramento multidimensional 45
- dimensões
 - em análise de correspondência 28
- distâncias
 - no Ajuste de escala multidimensional 41
 - no desdobramento multidimensional 46

E

- enrugar regressão
 - na Regressão Categórica 13
- Escala multidimensional 35, 37, 38
 - estatísticas 35
 - gráficos 35, 40, 41
 - model 38
 - opções 40
 - recursos adicionais do comando 42
 - restrições 39
 - saída 41
- escores do objeto
 - em análise de múltiplas correspondências 34
 - na análise de componentes principais para dados categóricos 20

- escores do objeto (*continuação*)
 - na Análise de correlação canônica não linear 25
- estatísticas de confiança
 - em análise de correspondência 29
- estatísticas descritivas
 - na Regressão Categórica 14

G

- gráfico de dispersão de ajuste
 - no desdobramento multidimensional 45
- gráfico de resíduos
 - no desdobramento multidimensional 45
- gráficos
 - em análise de correspondência 30
 - na Análise de correlação canônica não linear 25
 - na Regressão Categórica 15
 - no Ajuste de escala multidimensional 40, 41
- gráficos de carregamentos de componente
 - na análise de componentes principais para dados categóricos 22
- gráficos de categoria
 - em análise de múltiplas correspondências 35
 - na análise de componentes principais para dados categóricos 21
- gráficos de centroídes projetados
 - na análise de componentes principais para dados categóricos 21
- gráficos de correlação
 - no Ajuste de escala multidimensional 40
- gráficos de espaço comum
 - no Ajuste de escala multidimensional 40
 - no desdobramento multidimensional 45
- gráficos de espaço comum final
 - no desdobramento multidimensional 45
- gráficos de espaço comum inicial
 - no desdobramento multidimensional 45
- gráficos de espaços individuais
 - no Ajuste de escala multidimensional 40
 - no desdobramento multidimensional 45
- gráficos de estresse
 - no Ajuste de escala multidimensional 40
 - no desdobramento multidimensional 45
- gráficos de incícios múltiplos
 - no desdobramento multidimensional 45

gráficos de medidas de discriminação
em análise de múltiplas
correspondências 35

gráficos de ponderação de espaço
no desdobramento
multidimensional 45

gráficos de ponderação de espaço
individual
no Ajuste de escala
multidimensional 40

no desdobramento
multidimensional 45

gráficos de pontos de objeto
em análise de múltiplas
correspondências 35

na análise de componentes principais
para dados categóricos 21

gráficos de transformação
em análise de múltiplas
correspondências 35

na análise de componentes principais
para dados categóricos 21

no Ajuste de escala
multidimensional 40

no desdobramento
multidimensional 45

H

histórico de iteração
em análise de múltiplas
correspondências 34

na análise de componentes principais
para dados categóricos 20

no Ajuste de escala
multidimensional 41

no desdobramento
multidimensional 46

I

inclinação
em análise de múltiplas
correspondências 32

na análise de componentes principais
para dados categóricos 18

na Regressão Categórica 12

inércia
em análise de correspondência 29

J

juntar gráficos de categoria
em análise de múltiplas
correspondências 35

na análise de componentes principais
para dados categóricos 21

L

laço
na Regressão Categórica 13

M

matriz de correlações
em análise de múltiplas
correspondências 34

na análise de componentes principais
para dados categóricos 20

medidas de discriminação
em análise de múltiplas
correspondências 34

medidas de distância
em análise de correspondência 28

medidas de estresse
no Ajuste de escala
multidimensional 41

no desdobramento
multidimensional 46

modelo de ajuste de escala
no desdobramento
multidimensional 43

modelo Euclidiano generalizado
no desdobramento
multidimensional 43

modelo Euclidiano ponderado
no desdobramento
multidimensional 43

modelo identidade
no desdobramento
multidimensional 43

N

nível de ajuste de escala ideal
em análise de múltiplas
correspondências 32

na análise de componentes principais
para dados categóricos 17

normalização
em análise de correspondência 28

O

objetos complementares
na Regressão Categórica 13

P

padronização
em análise de correspondência 28

pesos
na Análise de correlação canônica não
linear 25

ponderação variável
em análise de múltiplas
correspondências 32

na análise de componentes principais
para dados categóricos 17

ponderações de espaço individual
no Ajuste de escala
multidimensional 41

no desdobramento
multidimensional 46

PREFSCAL 42

proximidades transformadas
no Ajuste de escala
multidimensional 41

proximidades transformadas (*continuação*)
no desdobramento
multidimensional 46

Q

quantificações de categoria
em análise de múltiplas
correspondências 34

na análise de componentes principais
para dados categóricos 20

na Análise de correlação canônica não
linear 25

na Regressão Categórica 14

R

R múltiplo
na Regressão Categórica 14

rede elástica
na Regressão Categórica 13

Regressão categórica 10

estatísticas 10

gráficos 10

nível de ajuste de escala ideal 11

recursos adicionais do comando 15

regularização 13

save 15

restrições
no Ajuste de escala
multidimensional 39

restrições em espaço comum
no desdobramento
multidimensional 44

T

termo de penalidade
no desdobramento
multidimensional 44

transformações de proximidade
no desdobramento
multidimensional 43

triplots
na análise de componentes principais
para dados categóricos 21

V

valores omissos
em análise de múltiplas
correspondências 32

na análise de componentes principais
para dados categóricos 18

na Regressão Categórica 12

variação apresentada
na análise de componentes principais
para dados categóricos 20

variáveis independentes transformadas
no Ajuste de escala
multidimensional 41



Impresso no Brasil