

*IBM SPSS Statistics Base 29*

**IBM**

**Nota**

Antes de usar estas informações e o produto suportado por elas, leia as informações nos [“Avisos”](#) na [página 277](#).

**Informações sobre o produto**

Essa edição se aplica à versão 29, liberação 0, modificação 1 de IBM® SPSS Statistics e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

© Copyright International Business Machines Corporation .

# Índice

<b>Capítulo 1. Recursos Principais.....</b>	<b>1</b>
Análise de potência.....	1
Médias.....	2
Proporções.....	16
Correlações.....	26
Regressão.....	35
Análise de energia: precisão.....	37
Análise de energia: valores de grade.....	38
Meta-análise.....	38
Metanálise Contínua.....	39
Tamanho do Efeito Contínuo de Metanálise.....	47
Binário de Metanálise.....	55
Tamanho do Efeito Binário de Metanálise.....	64
Regressão de Metanálise.....	73
Opções de plotagem de metanálise.....	77
Livro de códigos.....	83
Guia Saída do livro de códigos.....	84
Guia Estatísticas do livro de códigos.....	86
Frequências.....	86
Estatísticas de frequências.....	87
Gráficos de frequências.....	88
Formato de frequências.....	88
Descritivas.....	89
Opções Descritivas.....	89
Recursos adicionais do comando DESCRIPTIVES.....	90
Percentis.....	90
Valores omissos de percentis.....	91
Explorar.....	91
Explorar Estatísticas.....	92
Explorar Gráficos.....	92
Opções de exploração.....	93
Recursos adicionais do comando EXAMINE.....	93
Tabulações cruzadas.....	94
Camadas de Crisstabs.....	95
Gráfico de barras agrupadas de Crosstabs.....	95
Crosstabs exibindo variáveis de camada em camadas da tabela.....	95
Estatísticas de Crosstabs.....	95
Exibição da célula de tabulações cruzadas.....	97
Formato de tabela de Crosstabs.....	98
Resumir.....	98
Opções de sumarização.....	99
Estatísticas de sumarização.....	99
Médias.....	101
Opções de Médias.....	101
Cubos OLAP.....	103
Estatísticas de Cubos OLAP.....	104
Diferenças de Cubos OLAP.....	105
Título de Cubos OLAP.....	105
Proporções.....	106
Introdução de proporções.....	106
Proporções de uma amostra.....	106

Proporções de amostras pareadas.....	109
Proporções de amostras independentes.....	111
Testes t.....	114
Teste t.....	114
Teste T de amostras independentes.....	114
Teste-T de amostras em pares.....	116
Teste-T de uma amostra.....	117
Recursos adicionais do comando T TEST.....	119
ANOVA de um fator.....	119
Contrastes ANOVA de um fator.....	120
Testes posteriori de ANOVA de um fator.....	120
Opções de ANOVA unidirecional.....	122
Recursos adicionais do comando ONEWAY.....	123
Análise GLM Univariate.....	123
Modelo de GLM.....	125
Contrastes GLM.....	126
Gráficos de perfil GLM.....	127
Comparações a posteriori de GLM.....	129
Salvamento de GLM.....	132
Médias marginais estimadas de GLM.....	133
Opções de GLM.....	133
Recursos adicionais do comando UNIANOVA.....	134
Correlações bivariadas.....	135
Opções de correlações bivariadas.....	136
Intervalo de confiança de bivariate correlations.....	137
Recursos adicionais dos comandos CORRELATIONS e NONPAR CORR.....	137
Correlações parciais.....	137
Opções de correlações parciais.....	138
Recursos adicionais do comando PARTIAL CORR.....	138
Distâncias.....	139
Medidas de dissimilaridade de distâncias.....	139
Medidas de similaridade de distâncias.....	140
Recursos adicionais do comando PROXIMITIES.....	140
Modelos lineares.....	140
Para obter um modelo linear.....	141
Objetivos .....	141
Básicos .....	141
Seleção de Modelo .....	142
Combinações .....	143
Avançado .....	143
Opções de Modelo .....	143
Sumarização do Modelo .....	143
Preparação de Dado Automático .....	144
Importância do Preditor .....	144
Predito por Observado .....	144
Resíduos .....	144
Valores discrepantes .....	145
Efeitos .....	145
Coeficientes .....	145
Médias Estimadas .....	146
Sumarização de Construção de Modelo .....	146
Regressão linear.....	147
Métodos de Seleção de Variável de Regressão Linear.....	148
Regra do Conjunto de Regressão Linear.....	148
Gráficos de Regressão Linear.....	148
Regressão linear: salvando novas variáveis.....	149
Estatísticas de Regressão Linear.....	150
Opções de regressão linear.....	151

Recursos adicionais do comando REGRESSION.....	152
Regressão ordinal.....	152
Opções de Regressão Ordinal.....	153
Saída de Regressão Ordinal.....	153
Modelo de Localização de Regressão Ordinal.....	154
Modelo de Escala de Regressão Ordinal.....	155
Recursos Adicionais do Comando PLUM.....	155
Regressão de rede elástica linear.....	156
Regressão Líquida Elástica Lineares: Opções.....	156
Regressão linear lasso.....	159
Regressão Linear Lasso: Opções.....	160
Regressão de Ridge linear.....	162
Regressão Linear Ridge: Opções.....	163
Estimativa de curva.....	165
Modelos de Curva de Estimação.....	166
Salvar na Curva de Estimação.....	166
Regressão por quadrados mínimos parciais.....	167
Modelo .....	169
Opções .....	169
Análise do vizinho mais próximo.....	169
Vizinhos .....	171
Recursos .....	172
Partições .....	172
Salvar .....	173
Saída .....	174
Opções .....	174
Visualização do Modelo .....	174
Análise discriminante.....	177
Análise discriminante Definir intervalo.....	178
Casos Seleccionados de Análise Discriminante.....	178
Estatísticas de análise discriminante.....	179
Método de análise discriminante stepwise.....	179
Classificação da Análise Discriminante.....	180
Salvamento de análise discriminante.....	181
Recursos adicionais do comando DISCRIMINANT.....	181
Análise fatorial.....	181
Selecionar Casos para Análise Fatorial.....	182
Descritivas de Análise Fatorial.....	182
Extração de Análise Fatorial.....	182
Rotação de Análise Fatorial.....	183
Escore de Análise Fatorial.....	184
Opções de Análise Fatorial.....	184
Recursos Adicionais do Comando FACTOR.....	185
Escolhendo um Procedimento para Clusterização.....	185
Análise de cluster TwoStep.....	185
Opções de Análise de Cluster TwoStep.....	187
Saída da Análise de Cluster TwoStep.....	188
O visualizador de cluster.....	188
Análise de cluster hierárquica.....	193
Método de Análise de Cluster Hierárquica.....	194
Estatísticas de Análise de Cluster Hierárquica.....	194
Gráficos de Análise de Cluster Hierárquica.....	195
Análise de cluster hierárquica Salvar novas variáveis.....	195
Recursos Adicionais da Sintaxe do Comando CLUSTER.....	195
Análise de Cluster por K-Médias.....	195
Eficiência de Análise de Cluster por K-Médias.....	196
Iteração de Análise de Cluster por K-Médias.....	196
Salvar Análise de Cluster por K-Médias.....	197

Opções de análise de cluster por k-médias.....	197
Recursos adicionais do comando QUICK CLUSTER.....	197
Testes não paramétricos.....	198
Testes não paramétricos de amostra única.....	198
Testes Não Paramétricos de Amostras Independentes.....	202
Testes não paramétricos de amostras relacionadas.....	204
Recursos adicionais do comando NPTESTS.....	207
Diálogos Legado .....	207
Análise de Múltiplas Respostas.....	218
Análise de Múltiplas Respostas.....	218
Definir Conjuntos de Múltiplas Respostas.....	219
Frequências de múltiplas respostas.....	219
Tabulações cruzadas de múltiplas respostas.....	220
Resultados do relatório.....	222
Resultados do Relatório.....	222
Sumarizações de Relatórios em Linhas.....	222
Resumos de relatórios em colunas.....	224
Recursos adicionais do comando REPORT.....	226
Análise de Confiabilidade.....	227
Análise de confiabilidade: estatísticas.....	228
Recursos Adicionais do Comando RELIABILITY.....	230
Kappa ponderado.....	231
Kappa ponderado: Critérios.....	232
Kappa ponderado: imprimir.....	232
Ajuste de escala multidimensional.....	233
Forma de dados para ajuste de escala multidimensional.....	234
Criação de Medida para Ajuste de Escala Multidimensional.....	234
Modelo de Ajuste de Escala Multidimensional.....	234
Opções de ajuste de escala multidimensional.....	235
Recursos adicionais do comando ALSCAL.....	235
Estatísticas de razão.....	235
Estatísticas de razão: estatísticas.....	236
Estatísticas de razão.....	237
Estatísticas de razão: estatísticas.....	238
Gráfico P-P.....	239
Gráfico Q-Q.....	241
Análise de Característica de Operação do Receptor.....	244
Análise ROC: opções.....	245
Análise de Característica de Operação do Receptor: Exibição.....	246
Análise de ROC: Definir grupos (sequência).....	247
Análise de ROC: Definir grupos (numéricos).....	247
Curvas ROC.....	247
Opções de curva ROC.....	248
simulação.....	248
Para designar uma simulação baseada em um arquivo de modelo.....	249
Para designar uma simulação baseada em equações customizadas.....	249
Para designar uma simulação sem um modelo preditivo.....	250
Para executar uma simulação a partir de um plano de simulação.....	250
Construtor de simulação.....	251
Executar diálogo de simulação.....	262
Trabalhando com saída de gráfico a partir da simulação.....	264
Modelagem geoespacial.....	265
Selecionando Mapas .....	266
Origens de Dados .....	268
Regras de associação geoespacial .....	269
Predição temporal espacial .....	273
Conclusão .....	276

<b>Avisos.....</b>	<b>277</b>
Marcas comerciais.....	278
<b>Índice remissivo.....</b>	<b>281</b>





---

# Capítulo 1. Recursos Principais

Os recursos principais a seguir são incluídos no IBM SPSS Statistics Base Edition.

---

## Análise de potência

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

O IBM SPSS Statistics fornece os seguintes procedimentos de Análise de potência:

### Teste T de uma amostra

Na análise de uma amostra, os dados observados são coletados como uma única amostra aleatória. Assume-se que os dados amostrais seguem de forma independente e idêntica uma distribuição normal com média e variância fixas, e faz inferência estatística sobre o parâmetro médio.

### Teste T de amostra pareada

Na análise de amostra pareada, os dados observados contêm duas amostras pareadas e correlacionadas e cada caso possui duas medidas. Supõe-se que os dados em cada amostra seguem de forma independente e idêntica uma distribuição normal com média e variância fixas, e faz inferência estatística sobre a diferença das duas médias.

### Teste T de amostra independente

Na análise de amostra independente, os dados observados contêm duas amostras independentes. Supõe-se que os dados em cada amostra seguem de forma independente e idêntica uma distribuição normal com média e variância fixas, e faz inferência estatística sobre a diferença das duas médias.

### ANOVA de um fator

Análise de variância (ANOVA) é um método estatístico de estimar as médias de várias populações que, geralmente, são consideradas como distribuídas normalmente. A ANOVA de um fator, um tipo comum de ANOVA, é uma extensão do teste *t* de duas amostras.

**Exemplo.** A potência de um teste de hipótese para detectar uma hipótese alternativa correta é a probabilidade de que o teste rejeitará a hipótese de teste. Como a probabilidade de um erro tipo II é a probabilidade de aceitar a hipótese nula quando a hipótese alternativa é verdadeira, a potência pode ser expressa como  $(1 - \text{probabilidade de um erro tipo II})$ , que é a probabilidade de rejeitar a hipótese nula quando a hipótese alternativa é verdadeira.

**Estatísticas e gráficos.** Teste unilateral, teste bilateral, nível de significância, taxa de erro Tipo I, suposições de teste, desvio padrão de população, média populacional em teste, valor hipotético, potência bidimensional por tamanho da amostra, potência bidimensional por tamanho do efeito, potência tridimensional por tamanho da amostra, potência tridimensional por tamanho do efeito, graus de rotação, pares de grupos, coeficiente de correlação produto-momento de Pearson, diferença média, intervalo de gráfico do tamanho do efeito, desvio padrão agrupado da população, contrastes e diferenças de pairwise, coeficientes de contraste, teste de contraste, BONFERRONI, SIDAK, LSD, potência por tamanho total da amostra, potência bidimensional por desvio padrão agrupado, potência tridimensional por total da amostra, potência tridimensional por tamanho total da amostra, desvio padrão agrupado, distribuição t de Student, distribuição t não central,

## Considerações sobre dados da Análise de potência

### Dados

Na análise de uma amostra, os dados observados são coletados como uma única amostra aleatória.

Na análise de amostra pareada, os dados observados contêm duas amostras pareadas e correlacionadas e cada caso possui duas medidas.

Na análise de amostra independente, os dados observados contêm duas amostras independentes.

Análise de variância (ANOVA) é um método estatístico de estimar as médias de várias populações que, geralmente, são consideradas como distribuídas normalmente.

### Suposições

Na análise de uma amostra, considera-se que os dados de amostra sigam, de forma independente e idêntica, uma distribuição normal com uma média e variância fixas, e façam inferência estatística sobre o parâmetro médio.

Na análise de amostra pareada, supõe-se que os dados sigam, de forma independente e idêntica, uma distribuição normal com uma média e variância fixas, e façam inferência estatística sobre a diferença das duas médias.

Na análise de amostra independente, supõe-se que os dados sigam, de forma independente e idêntica, uma distribuição normal com uma média e variância fixas, e façam inferência estatística sobre a diferença das duas médias.

Na ANOVA de um fator, o método estatístico de estimar as médias de várias populações que geralmente são consideradas como distribuídas normalmente.

## Obtendo uma Análise de potência

1. A partir dos menus, escolha:

**Analisar > Análise de potência > Comparar médias > Teste T de uma amostra**, ou **Teste T de amostra pareada**, ou **Teste T de amostra independente** ou **ANOVA de um fator**

2. Defina as suposições de teste necessárias.

3. Clique em **OK**.

## Médias

Os seguintes recursos estatísticos estão incluídos no IBM SPSS Statistics Base Edition.

### Análise de potência do teste T de uma amostra

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

Na análise de uma amostra, os dados observados são coletados como uma única amostra aleatória. Assume-se que os dados amostrais seguem de forma independente e idêntica uma distribuição normal com média e variância fixas, e faz inferência estatística sobre o parâmetro médio.

1. A partir dos menus, escolha:

**Analisar > Análise de potência > Médias > Teste T de uma amostra**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.

3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione **Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade”](#) na página 38.

4. Quando **Potência** for selecionada como o método de **estimativa** de suposição de teste, insira um **tamanho de amostra** apropriado para o valor de estimativa de potência. O valor deve ser um número inteiro maior que 1.
5. Opcionalmente, selecione uma opção no campo **Especificar**.

#### **Valores hipotéticos**

A configuração padrão fornece as configurações de **média populacional** e **valor nulo**.

##### **Média populacional**

Insira um valor que especifique a média populacional em teste. O valor deve ser um valor numérico único.

##### **Valor nulo**

Opcionalmente, insira um valor que especifique o valor da hipótese nula a ser testado. O valor deve ser um valor numérico único.

#### **Tamanho do efeito**

O  $f$  de Cohen é usado para estimar o tamanho do efeito como uma entrada para a estimativa do poder ou tamanho da amostra. O **valor** do tamanho de efeito definido é passado para a etapa intermediária no procedimento e calcula o poder desejado ou o tamanho da amostra.

6. Insira um valor de **Desvio padrão de população**. O valor deve ser um valor numérico único maior que 0.
7. Selecione se o teste é unilateral ou bilateral.

##### **Análise não direcional (bilateral)**

Quando selecionada, é usado um teste bilateral. Essa é a configuração padrão.

##### **Análise direcional (unilateral)**

Quando selecionada, a potência é calculada para um teste unilateral.

8. Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
9. Opcionalmente, você pode clicar em **Criar gráfico** para especificar as configurações de “Análise de potência de teste T de uma amostra: Gráfico” na página 3 (saída do gráfico, configurações de plotagem bidimensional e configurações de plotagem tridimensional).

**Nota:** O **Criar gráfico** está disponível somente quando **Potência** é selecionada como a suposição de teste.

10. Opcionalmente, clique em **Precisão** para estimar o tamanho da amostra com base em intervalos de confiança, especificando os valores médios do intervalo de confiança. Para obter mais informações, consulte “Análise de energia: precisão” na página 37.

**Nota:** A **Precisão** está disponível somente quando **Tamanho da amostra** é selecionado como o método de **Estimativa** de suposição de teste, **Valores hipotéticos** são selecionados na lista **Especificar** e **Análise não direcional (frente e verso)** é selecionada como **Direção de teste**.

### **Análise de potência de teste T de uma amostra: Gráfico**

A caixa de diálogo **Criar gráfico** fornece opções para controlar os gráficos que são produzidos para ilustrar o poder bidimensional e tridimensional por gráficos de tamanho de amostra e efeito. A caixa de diálogo também controla os graus de rotação vertical e horizontal para gráficos tridimensionais.

#### **Gráfico bidimensional**

##### **Estimativa de potência versus tamanho da amostra**

Quando ativada, esta configuração opcional oferece opções para controlar o poder bidimensional por gráfico de tamanho de amostra. Este serviço está desativado por padrão.

##### **Intervalo de tamanho da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho da amostra será usado.

**Limite inferior**

Controla o limite inferior para a potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior que 1 e não pode ser maior que o valor de **Limite superior**.

**Limite superior**

Controla o limite superior para a potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

**Estimação de potência versus tamanho do efeito**

Por padrão, esta configuração opcional é desativada. Quando ativada, o gráfico é exibido na saída. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho do efeito será usado.

**Intervalo de tamanho do efeito**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis.

**Limite inferior**

Controla o limite inferior para a potência bidimensional por gráfico de tamanho do efeito. O valor deve ser maior ou igual a -5,0 e não pode ser maior que o valor de **Limite superior**.

**Limite superior**

Controla o limite superior para a potência bidimensional por gráfico de tamanho do efeito. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5,0.

**Gráfico tridimensional****Estimação de potência versus**

Fornecer opções para controlar a potência tridimensional por tamanho de amostra (eixo x) e gráfico de tamanho de efeito (eixo y), as configurações de rotação vertical/horizontal e o intervalo de plotagem especificado pelo usuário do tamanho de amostra/efeito. Esta configuração é desativada por padrão.

**Tamanho do efeito no eixo X e tamanho da amostra no eixo Y**

A configuração opcional controla o poder tridimensional pelo gráfico de tamanho da amostra (eixo x) e tamanho do efeito (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

**Tamanho do efeito no eixo Y e tamanho da amostra no eixo X**

A configuração opcional controla a potência tridimensional pelo gráfico de tamanho da amostra (eixo y) e tamanho do efeito (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

**Rotação vertical**

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

**Rotação horizontal**

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

**Intervalo de tamanho da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho da amostra será usado.

**Limite inferior**

Controla o limite inferior para a potência tridimensional por gráfico de tamanho da amostra. O valor deve ser maior que 1 e não pode ser maior que o valor de **Limite superior**.

### **Limite superior**

Controla o limite superior para a potência tridimensional por gráfico de tamanho da amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

### **Intervalo de tamanho do efeito**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho do efeito será usado.

### **Limite inferior**

Controla o limite inferior para a potência tridimensional por gráfico de tamanho do efeito. O valor deve ser maior ou igual a -5,0 e não pode ser maior que o valor de **Limite superior**.

### **Limite superior**

Controla o limite superior para a potência tridimensional por gráfico de tamanho do efeito. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5,0.

## **Análise de potência do teste T de amostras pareadas**

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

Na análise de amostras pareadas, os dados observados contêm duas amostras pareadas e correlacionadas, e cada caso tem duas medidas. Assume-se que os dados em cada amostra seguem de forma independente e idêntica uma distribuição normal com média e variância fixas, e faz inferência estatística sobre a diferença das duas médias.

1. A partir dos menus, escolha:

### **Analisar > Análise de potência > Médias > Teste T de Amostras Pareadas**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.
3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione **Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade”](#) na página 38.

4. Quando **Potência** for selecionado como o método de **estimativa** de suposição de teste, insira o valor de **tamanho de amostra** apropriado. O valor deve ser um número inteiro positivo maior ou igual a 2.
5. Opcionalmente, selecione uma opção a partir do campo **Especificar**.

### **Valores hipotéticos**

A configuração padrão fornece as configurações de **Média** e **Desvio Padrão**.

#### **Diferença de média populacional**

Quando uma única média populacional for necessária, insira um valor de **Diferença média populacional**. Quando um valor único é especificado, ele indica a diferença média populacional  $\mu_d$ .

**Nota:** O valor não pode ser 0 quando o **tamanho da amostra** é selecionado.

#### **Média populacional para o grupo 1 e grupo 2**

Quando várias médias populacionais forem necessárias para os pares de grupos especificados, insira valores para **Média populacional para o grupo 1** e **Média populacional para o grupo 2**.

Quando vários valores são especificados, eles denotam a diferença média da população  $\mu_1$  e  $\mu_2$ .

**Nota:** Os dois valores não podem ser iguais quando o **tamanho da amostra** é selecionado.

#### **Desvio padrão da população para diferença média**

Quando uma única média populacional for especificada, insira o desvio padrão da população para o valor da diferença média. Quando um valor único é especificado, ele indica o desvio padrão populacional da diferença de grupo  $\sigma_d$ . O valor deve ser um valor numérico único maior que 0.

#### **Desvio padrão da população para o grupo 1 e grupo 2**

Quando várias médias de população são especificadas, insira o desvio padrão da população para os valores do grupo 1 e do grupo 2. Quando diversos valores são especificados, eles indicam o desvio padrão populacional da diferença de grupo  $\sigma_1$  e  $\sigma_2$ . Os valores devem ser um único numérico maior que 0.

#### **Coefficiente de correlação do momento do produto da Pearson**

Opcionalmente, insira um valor que especifique o coeficiente de correlação do momento do produto da Pearson  $\rho$ . O valor deve ser um único valor numérico entre -1 e 1. O valor não pode ser 0.

**Nota:** Quando um valor único de **Desvio padrão de população para diferença média** é especificado, esta configuração é ignorada. Caso contrário, os valores de **desvio padrão da população para o grupo 1 e grupo 2** são usados para calcular  $\sigma_d$ .

#### **Tamanho do efeito**

Estima o tamanho do efeito como uma entrada para a estimativa do poder ou tamanho da amostra. O **valor** do tamanho de efeito definido é passado para a etapa intermediária no procedimento e calcula o poder desejado ou o tamanho da amostra.

6. Selecione se o teste é unilateral ou bilateral.

#### **Análise não direcional (bilateral)**

Quando selecionada, é usado um teste bilateral. Essa é a configuração padrão.

#### **Análise direcional (unilateral)**

Quando selecionada, a potência é calculada para um teste unilateral.

7. Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
8. Opcionalmente, clique em **Criar gráfico** para especificar as configurações de “[Análise de potência de teste T de amostras pareadas: Gráfico](#)” na [página 6](#) (saída do gráfico, configurações de plotagem bidimensional, configurações de plotagem tridimensional e dicas de ferramentas).

**Nota:** O **Criar gráfico** está disponível somente quando **Potência** é selecionada como a suposição de teste.

9. Opcionalmente, clique em **Precisão** para estimar o tamanho da amostra com base em intervalos de confiança, especificando os valores médios do intervalo de confiança. Para obter mais informações, consulte “[Análise de energia: precisão](#)” na [página 37](#).

**Nota:** **Precisão** está disponível apenas quando **Tamanho da amostra** é selecionado como o método de **Estimativa** de teste e **Valores hipotéticos** são selecionados a partir da lista **Especificar**.

### ***Análise de potência de teste T de amostras pareadas: Gráfico***

A caixa de diálogo **Criar gráfico** fornece opções para controlar os gráficos que são produzidos para ilustrar o poder bidimensional e tridimensional por gráficos de tamanho de amostra e efeito. A caixa de diálogo também controla os graus de rotação vertical e horizontal para gráficos tridimensionais.

#### **Gráfico bidimensional**

##### **Estimativa de potência versus tamanho da amostra**

Quando ativada, esta configuração opcional oferece opções para controlar o poder bidimensional por gráfico de tamanho de amostra. Este serviço está desativado por padrão.

### **Intervalo de tamanho da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho da amostra será usado.

#### **Limite inferior**

Controla o limite inferior para a potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior que 1 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para a potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

### **Estimação de potência versus tamanho do efeito**

Por padrão, esta configuração opcional é desativada. Quando ativada, o gráfico é exibido na saída. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho do efeito será usado.

#### **Intervalo de tamanho do efeito**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis.

##### **Limite inferior**

Controla o limite inferior para a potência bidimensional por gráfico de tamanho do efeito. O valor deve ser maior ou igual a -5,0 e não pode ser maior que o valor de **Limite superior**.

##### **Limite superior**

Controla o limite superior para a potência bidimensional por gráfico de tamanho do efeito. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5,0.

### **Gráfico tridimensional**

#### **Estimação de potência versus**

Fornecer opções para controlar a potência tridimensional por tamanho de amostra (eixo x) e gráfico de tamanho de efeito (eixo y), as configurações de rotação vertical/horizontal e o intervalo de plotagem especificado pelo usuário do tamanho de amostra/efeito. Esta configuração é desativada por padrão.

#### **Tamanho do efeito no eixo X e tamanho da amostra no eixo Y**

A configuração opcional controla o poder tridimensional pelo gráfico de tamanho da amostra (eixo x) e tamanho do efeito (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

#### **Tamanho do efeito no eixo Y e tamanho da amostra no eixo X**

A configuração opcional controla a potência tridimensional pelo gráfico de tamanho da amostra (eixo y) e tamanho do efeito (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

#### **Rotação vertical**

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

#### **Rotação horizontal**

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

#### **Intervalo de tamanho da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho da amostra será usado.

**Limite inferior**

Controla o limite inferior para a potência tridimensional por gráfico de tamanho da amostra. O valor deve ser maior que 1 e não pode ser maior que o valor de **Limite superior**.

**Limite superior**

Controla o limite superior para a potência tridimensional por gráfico de tamanho da amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

**Intervalo de tamanho do efeito**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho do efeito será usado.

**Limite inferior**

Controla o limite inferior para a potência tridimensional por gráfico de tamanho do efeito. O valor deve ser maior ou igual a -5,0 e não pode ser maior que o valor de **Limite superior**.

**Limite superior**

Controla o limite superior para a potência tridimensional por gráfico de tamanho do efeito. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5,0.

## Análise de potência do teste T de amostras independentes

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

Na análise de amostras independentes, os dados observados contêm duas amostras independentes. Assume-se que os dados em cada amostra seguem de forma independente e idêntica uma distribuição normal com média e variância fixas, e faz inferência estatística sobre a diferença das duas médias.

1. A partir dos menus, escolha:

**Analisar > Análise de potência > Médias > Teste T de Amostras Independentes**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.
3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione **Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade” na página 38](#).

Insira um valor **Razão de tamanho do grupo** para especificar a razão dos tamanhos da amostra (o valor deve ser um valor único entre 0,01 e 100).

4. Quando **Potência** for selecionado como o método de **Estimativa** de suposição de teste, insira valores para especificar o tamanho da amostra para **Tamanho da amostra para o grupo 1 e o grupo 2** para comparação. Os valores devem ser números inteiros maiores que 1.
5. Opcionalmente, selecione uma opção no campo **Especificar**.

**Valores hipotéticos**

A configuração padrão fornece as configurações de **Média populacional** e **Desvios padrão populacionais**.



### Diferença média populacional

Quando uma única média populacional for especificada, insira o valor da diferença média populacional. Quando um único valor é especificado, denota a população da diferença média do grupo  $\sigma_d$ . O valor deve ser um valor numérico único maior que 0.

### Média populacional para o grupo 1 e grupo 2

Quando várias médias populacionais forem especificadas, insira a média populacional para os valores do grupo 1 e do grupo 2. Quando vários valores são especificados, eles denotam a média populacional da diferença de grupo  $\sigma_1$  e  $\sigma_2$ . Os valores devem ser um único numérico maior que 0.

### Desvios padrão da população são

Especifique se os desvios padrão de população são **Iguais para dois grupos** ou **Diferentes para dois grupos**.

- Quando os desvios padrão de população forem iguais para dois grupos, insira um valor para **Desvio padrão agrupado** que denota  $\sigma$  e considera que as duas variâncias de grupos são iguais ou  $\sigma_1 = \sigma_2 = \sigma$ .
- Quando os desvios padrão da população não são iguais para dois grupos, insira valores para **Desvio padrão para o grupo 1 e grupo 2** que denotam  $\sigma_1$  e  $\sigma_2$ .

**Nota:** Quando os valores de **Desvio padrão para o grupo 1 e grupo 2** são idênticos, eles são tratados como um único valor.

### Tamanho do efeito

Estima o tamanho do efeito como uma entrada para a estimativa do poder ou tamanho da amostra. O **valor** do tamanho de efeito definido é passado para a etapa intermediária no procedimento e calcula o poder desejado ou o tamanho da amostra.

6. Selecione se o teste é unilateral ou bilateral.

#### Análise não direcional (bilateral)

Quando selecionada, é usado um teste bilateral. Essa é a configuração padrão.

#### Análise direcional (unilateral)

Quando selecionada, a potência é calculada para um teste unilateral.

7. Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
8. Opcionalmente, clique em **Criar gráfico** para especificar as configurações de “Análise de potência de teste T de amostras independentes: Gráfico” na página 9 (saída do gráfico, configurações de plotagem bidimensional, configurações de plotagem tridimensional e dicas de ferramentas).

**Nota:** O **Criar gráfico** está disponível somente quando **Potência** é selecionada como a suposição de teste.

9. Opcionalmente, clique em **Precisão** para estimar o tamanho da amostra com base em intervalos de confiança, especificando os valores médios do intervalo de confiança. Para obter mais informações, consulte “Análise de energia: precisão” na página 37.

**Nota:** A **Precisão** está disponível somente quando **Tamanho da amostra** é selecionado como o método de **Estimativa** de suposição de teste, **Valores hipotéticos** são selecionados na lista **Especificar** e **Análise não direcional (frente e verso)** é selecionada como **Direção de teste**.

### Análise de potência de teste T de amostras independentes: Gráfico

A caixa de diálogo **Criar gráfico** oferece opções para controlar os gráficos que são emitidos para ilustrar a potência bidimensional e tridimensional por razão de amostra, tamanho do efeito ou gráficos de diferença média. A caixa de diálogo também controla os graus de rotação vertical e horizontal para gráficos tridimensionais.

## Gráfico bidimensional

### **Estimação de energia versus razão de tamanho da amostra/Estimação de energia versus tamanho da amostra**

Quando ativada, esta configuração opcional oferece opções para controlar o gráfico de proporção de potência bidimensional por tamanho de amostra. As configurações adicionais não estarão disponíveis quando vários valores de energia forem especificados (**Estimação de energia versus tamanho da amostra**). Este serviço está desativado por padrão.

#### **Intervalo de razão de tamanho da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho da amostra será usado.

#### **Limite inferior**

Controla o limite inferior para a potência bidimensional por gráfico de razão de tamanho da amostra. O valor deve estar entre 0,01 e 100 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para a potência bidimensional por gráfico de razão de tamanho da amostra. O valor deve estar entre 0,01 e 100 e deve ser maior que o valor de **Limite inferior**.

### **Estimativa de potência versus tamanho do efeito (ou diferença média)**

Por padrão, esta configuração opcional é desativada. Quando ativada, o gráfico é exibido na saída. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho do efeito (ou diferença média) será usado.

#### **Intervalo de tamanho do efeito (ou diferença média)**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis.

#### **Limite inferior**

Controla o limite inferior para a potência bidimensional por gráfico de tamanho do efeito. O valor deve ser maior ou igual a -5,0 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para a potência bidimensional por gráfico de tamanho do efeito. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5,0.

## Gráfico tridimensional

### **Estimação de potência versus**

Fornece opções para controlar a potência tridimensional por proporção de tamanho de amostra (eixo x) e gráfico de tamanho de efeito (eixo y), as configurações de rotação vertical/horizontal e o intervalo de plotagem especificado pelo usuário de tamanho de amostra/efeito. Esta configuração é desativada por padrão.

#### **Tamanho do efeito (ou diferença média) no eixo X e tamanho da amostra no eixo Y**

A configuração opcional controla a potência tridimensional por gráfico de proporção de tamanho de amostra (eixo x) e gráfico de tamanho de efeito (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

#### **Tamanho do efeito (ou diferença média) no eixo Y e tamanho da amostra no eixo X**

A configuração opcional controla a potência tridimensional por gráfico de tamanho da amostra (eixo y) e por gráfico de tamanho do efeito (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

#### **Rotação vertical**

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

### **Rotação horizontal**

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

### **Intervalo de razão de tamanho da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho da amostra será usado.

#### **Limite inferior**

Controla o limite inferior para a potência tridimensional por gráfico de razão de tamanho da amostra. O valor deve estar entre 0,01 e 100 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para a potência tridimensional por gráfico de razão de tamanho da amostra. O valor deve estar entre 0,01 e 100 e deve ser maior que o valor de **Limite inferior**.

### **Intervalo de tamanho do efeito (ou diferença média)**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho do efeito será usado.

#### **Limite inferior**

Controla o limite inferior para a potência tridimensional por gráfico de tamanho do efeito. O valor deve ser maior ou igual a -5,0 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para a potência tridimensional por gráfico de tamanho do efeito. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

## **Análise de potência de ANOVA de um fator**

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

Análise de variância (ANOVA) é um método estatístico de estimar as médias de várias populações que, geralmente, são consideradas como distribuídas normalmente. A ANOVA de um fator, um tipo comum de ANOVA, é uma extensão do teste *t* de duas amostras. O procedimento fornece abordagens para estimar a potência para dois tipos de hipótese para comparar as várias médias de grupo, o teste geral e o teste com contrastes especificados. O teste geral concentra-se na hipótese nula de que todas as médias de grupo são iguais. O teste com contrastes especificados divide as hipóteses de ANOVA gerais em partes de médias menores, porém mais descritíveis e úteis.

1. A partir dos menus, escolha:

**Analisar > Análise de potência > Médias > One-way ANOVA**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.

3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione **Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade”](#) na página 38.

4. Opcionalmente, selecione uma opção no campo **Especificar**.

#### **Valores hipotéticos**

A configuração padrão fornece a configuração **Desvio padrão populacional agrupado**. Insira um valor numérico único maior que 0. O valor padrão é 1.

#### **Tamanho do efeito: Cohen's f**

O *f* de Cohen é usado para estimar o tamanho do efeito como uma entrada para a estimativa da potência ou do tamanho da amostra. O **valor** do tamanho do efeito definido é passado para a etapa intermediária no procedimento e calcula a potência ou o tamanho da amostra desejado.

#### **Tamanho do efeito: Eta-squared**

O Eta quadrado ( $\eta^2$ ) é usado para estimar o tamanho do efeito como uma entrada para a estimativa da potência ou do tamanho da amostra. O **valor** do tamanho do efeito definido é passado para a etapa intermediária no procedimento e calcula a potência ou o tamanho da amostra desejado.

5. Quando **Potência** é selecionado como a configuração de suposição de teste **Estimativa**, especifique os valores de **Tamanhos do grupo** e **Médias do grupo**. Pelo menos dois valores de tamanho do grupo devem ser especificados, e cada um dos valores de tamanho do grupo deve ser maior ou igual a 2 (mas não pode exceder 2.147. 483. 647). Pelo menos dois valores médios de grupo também devem ser especificados (o número de valores especificados deve ser igual aos valores do tamanho do grupo).

Quando **Tamanho da amostra** é selecionado como a configuração de suposição de teste **Estimativa**, especifique os valores de **Pesos do grupo** (opcional) e **Médias do grupo**. Os pesos do grupo designam os pesos do tamanho do grupo quando **Potência** é selecionada.

**Nota:** As configurações de **Ponderações de grupo** são ignoradas quando valores de **Tamanhos do grupo** são especificados.

Clique em **Incluir** para incluir mais valores de **Tamanhos do grupo**, **Pesos do grupo** e **Médias do grupo**. Clique em **Excluir** para remover os valores existentes.

6. Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
7. Opcionalmente, clique em **Contraste** para especificar configurações de [“Análise de potência de ANOVA de um fator: Contraste”](#) na página 12 (teste de contraste, diferenças pairwise, tamanhos de efeito e estimar o tamanho da mostra com base em intervalos de confiança).
8. Opcionalmente, clique em **Criar gráfico** para especificar as configurações de [“Análise de potência de ANOVA de um Fator: Gráfico”](#) na página 13 (saída do gráfico, configurações de plotagem bidimensional, configurações de plotagem tridimensional e dicas de ferramentas).

**Nota:** **Criar gráfico** está disponível apenas quando os valores de **Tamanhos do grupo** são especificados e **Potência** é selecionada como a configuração de **Estimativa**.

9. Opcionalmente, clique em **Precisão** para estimar o tamanho da amostra com base em intervalos de confiança, especificando os valores médios do intervalo de confiança. Para obter mais informações, consulte [“Análise de energia: precisão”](#) na página 37.

**Nota:** **Precisão** está disponível apenas quando **Tamanho da amostra** é selecionado como o método de suposição de teste **Estimativa**.

### ***Análise de potência de ANOVA de um fator: Contraste***

A caixa de diálogo **Contraste** fornece opções para especificar contraste, coeficiente, tamanho do efeito, diferenças de pares e configurações de meia largura do intervalo de confiança para a Análise de energia da análise ANOVA de um fator.

#### **Teste de contraste**

##### **Teste com contrastes lineares**

Quando ativada, as configurações de contraste e coeficiente estão disponíveis.

## Direção de teste

### Análise não direcional (bilateral)

Quando selecionada, é usado um teste bilateral. Essa é a configuração padrão.

### Análise direcional (unilateral)

Quando selecionada, a potência é calculada para um teste unilateral.

## Coeficientes

Use a tabela para especificar os coeficientes de contraste e solicitar o teste de contraste. Os valores da tabela são opcionais. O número de valores especificados deve ser igual aos valores especificados para **Tamanhos do grupo** e **Médias do grupo**. A soma de todos os valores especificados deve ser igual a 0, caso contrário, o último valor será ajustado automaticamente.

## Tamanho do efeito

Especifique o tamanho do efeito do teste de contraste que é medido por  $\delta$ . Deve-se especificar um único valor numérico. Quando a **Direção do teste** é configurada para **Análise direcional (unilateral)**, o valor especificado deve ser  $\geq 0$ . Quando a **Potência** é especificada como a suposição de teste **Estimativa**, o valor especificado não pode ser 0.

## Diferenças de pares

### Estime a potência do teste para diferenças de pares

Controla se estimar ou não a potência de teste para as diferenças pairwise. Por padrão, a configuração opcional é desabilitada, o que suprime a saída para as diferenças de pares.

### Correção de Bonferroni

Usa a correção de Bonferroni na estimação da potência de diferenças pairwise. Essa é a configuração padrão.

### correção de Sidak

Usa a correção de Sidak na estimação da potência de diferenças pairwise.

### Diferença menos significativa (LSD)

Usa a correção de LSD na estimação da potência de diferenças pairwise.

## Especificar a metade da largura do intervalo de confiança

Estima o tamanho da amostra com base no valor de meia largura do intervalo de confiança. Insira um valor no intervalo de 0 a 1. Para o teste Binomial de Uma Amostra, o valor deve estar no intervalo de 0 a 0,5.

**Nota:** Valores duplicados são ignorados.

- Clique em **Incluir** para incluir o valor de meia largura especificado na lista.
- Realce um valor de meia largura existente e clique em **Mudar** para atualizar o valor.
- Destaque um valor de meia largura existente e clique em **Remover** para remover o valor da lista.

## Análise de potência de ANOVA de um Fator: Gráfico

A caixa de diálogo **Criar gráfico** controla os gráficos que são produzidos para ilustrar o poder bidimensional e tridimensional por gráficos de tamanho de amostra e efeito. A caixa de diálogo também controla os graus de rotação vertical e horizontal para gráficos tridimensionais.

### Gráfico bidimensional

#### Estimativa de potência versus tamanho total da amostra

Quando ativada, esta configuração opcional oferece opções para controlar a potência bidimensional pelo gráfico de tamanho total da amostra. Este serviço está desativado por padrão.

#### Intervalo de tamanho total da amostra

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho total da amostra será usado.

**Limite inferior**

Controla o limite inferior para a potência bidimensional por gráfico de tamanho total da amostra. O valor deve ser maior ou igual a:

- 2 x a quantidade de números inteiros especificados para **Tamanhos de grupos**
- 2 x a soma dos números inteiros especificados para **Tamanhos de grupos** / pelo menor valor de número inteiro para **Tamanhos de grupos**

O valor não pode ser maior que o valor de **Limite superior**.

**Limite superior**

Controla o limite superior para a potência bidimensional por gráfico de tamanho total da amostra. O valor deve ser menor ou igual a:

- 5000 / pelo maior valor de número inteiro especificado para **Tamanhos de grupos** x a soma dos números inteiros especificados para **Tamanhos de grupos**

O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 2147483647.

**Estimativa de potência versus desvio padrão agrupado**

Por padrão, esta configuração opcional é desativada. A configuração controla a potência bidimensional por gráfico de desvio padrão agrupado. Quando ativada, o gráfico é exibido na saída. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do desvio padrão agrupado será usado.

**Nota:**

O plot é desativado quando os valores especificados **Médias do grupo** são todos iguais.

**Intervalo de desvio padrão agrupado**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis.

**Limite inferior**

Controla o limite inferior para a potência bidimensional por gráfico de desvio padrão agrupado. O valor deve ser maior que 0 e não pode ser maior que o valor de **Limite superior**.

**Limite superior**

Controla o limite superior para a potência bidimensional por gráfico de desvio padrão agrupado. O valor deve ser maior que o valor de **Limite inferior**.

**Gráfico tridimensional****Estimação de potência versus**

Fornece opções para controlar a potência tridimensional por gráfico de tamanho total da amostra (eixo x) e tamanho do efeito (eixo y), as configurações de rotação vertical e horizontal e o intervalo de plotagem especificado pelo usuário da amostra e tamanho do efeito. Esta configuração é desativada por padrão.

**Nota:**

O plot é desativado quando os valores especificados **Médias do grupo** são todos iguais.

**Desvio padrão agrupado no eixo X e tamanho total da amostra no eixo Y**

A configuração opcional controla a potência tridimensional por gráfico de tamanho total da amostra (eixo x) e desvio padrão agrupado (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

**Desvio padrão agrupado no eixo Y e tamanho total da amostra no eixo X**

A configuração opcional controla a potência tridimensional por gráfico de tamanho total da amostra (eixo y) e desvio padrão agrupado (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

### **Intervalo de tamanho total da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho da amostra será usado.

#### **Limite inferior**

Controla o limite inferior para a potência tridimensional por gráfico de tamanho total da amostra. O valor deve ser maior que 0 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para a potência tridimensional por gráfico de tamanho total da amostra. O valor deve ser maior que o valor de **Limite inferior**.

### **Intervalo de desvio padrão agrupado**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho do efeito será usado.

#### **Limite inferior**

Controla o limite inferior para a potência tridimensional por gráfico de desvio padrão agrupado. O valor deve ser maior que 0 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para a potência tridimensional por gráfico de desvio padrão agrupado. O valor deve ser maior que o valor de **Limite inferior**.

### **Estimativa de potência versus tamanho da amostra**

Fornece opções para controlar a potência tridimensional por tamanho de efeito e o intervalo de plotagem especificado pelo usuário de amostra e tamanho de efeito. Esta configuração é desativada por padrão.

#### **Nota:**

O plot é desativado quando os valores especificados **Médias do grupo** são todos iguais.

#### **Desvio padrão agrupado no eixo X e tamanho total da amostra no eixo Y**

A configuração opcional controla a potência tridimensional por gráfico de tamanho total da amostra (eixo x) e tamanho do efeito (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

#### **Desvio padrão agrupado no eixo Y e tamanho total da amostra no eixo X**

A configuração opcional controla a potência tridimensional por gráfico de tamanho total da amostra (eixo y) e tamanho do efeito (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

### **Intervalo de tamanho total da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho da amostra será usado.

#### **Limite inferior**

Controla o limite inferior para a potência tridimensional por gráfico de tamanho total da amostra. O valor deve ser maior que 0 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para a potência tridimensional por gráfico de tamanho total da amostra. O valor deve ser maior que o valor de **Limite inferior**.

### **Intervalo de tamanho do efeito**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão do tamanho do efeito será usado.

**Limite inferior**

Controla o limite inferior para a potência tridimensional por gráfico de tamanho total da amostra. O valor deve ser maior que 0 e não pode ser maior que o valor de **Limite superior**.

**Limite superior**

Controla o limite superior para a potência tridimensional por gráfico de tamanho total da amostra. O valor deve ser maior que o valor de **Limite inferior**.

**Rotação vertical**

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

**Rotação horizontal**

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

## Proporções

Os seguintes recursos estatísticos estão incluídos no IBM SPSS Statistics Base Edition.

### Análise de potência do teste binomial de uma amostra

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

A distribuição binomial é baseada em uma sequência de avaliações de Bernoulli. Ele pode ser usado para modelar experimentos, incluindo um número fixo de tentativas totais, que são consideradas independentes umas das outras. Cada tentativa leva a um resultado dicotômico, com a mesma probabilidade de sucesso.

O teste binomial de uma amostra faz inferência estatística sobre o parâmetro de proporção comparando-o com um valor hipotético. Os métodos para estimar o poder de tal teste são a aproximação normal ou a enumeração binomial.

1. A partir dos menus, escolha:

**Analisar > Análise de potência > Proporções > Teste binomial de uma amostra**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.
3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione **Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade”](#) na página 38.

4. Quando **Potência** for selecionada como a configuração de **Estimativa** da suposição de teste, insira o valor de **Número total de tentativas apropriado**. O valor deve ser um número inteiro maior ou igual a 1.
5. Insira um valor que especifica o valor de hipótese alternativa do parâmetro de proporção no campo **Proporção da população**. O valor deve ser um valor numérico único.



**Nota:** Quando um valor de **Potência** é especificado, o valor da **Proporção de população** não pode ser igual ao **valor nulo**.

6. Opcionalmente, insira um valor que especifica o valor de hipótese nula do parâmetro de proporção a ser testado no campo **Valor nulo**. O valor deve ser um único numérico no intervalo de 0 a 1. O valor padrão é 0,50.
7. Selecione um método para estimar a potência.

#### **Aproximação normal**

Ativa a aproximação normal. Essa é a configuração padrão.

#### **Aplicar correção de continuidade**

Controle se a correção de continuidade é usada ou não para o método de aproximação normal.

#### **Enumeração binomial**

Ativa o método de enumeração binomial. Opcionalmente, use o campo **Limite de tempo** para especificar o número máximo de minutos permitidos para estimar o tamanho da amostra. Quando o limite de tempo é atingido, a análise é finalizada e uma mensagem de aviso é exibida. Quando especificado, o valor deve ser um único inteiro positivo para denotar o número de minutos. A configuração padrão é de 5 minutos.

**Nota:** A suposição de estimativa de potência selecionada não tem efeito quando o valor do **número total de tentativas** excede 500.

8. Selecione se o teste é unilateral ou bilateral.

#### **Análise não direcional (bilateral)**

Quando selecionada, é usado um teste bilateral. Essa é a configuração padrão.

#### **Análise direcional (unilateral)**

Quando selecionada, a potência é calculada para um teste unilateral.

9. Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
10. Opcionalmente, clique em **Criar gráfico** para especificar as configurações de “Análise de potência de binômio de uma amostra: Gráfico” na página 17 (saída do gráfico, configurações de plotagem bidimensional e configurações de plotagem tridimensional).

**Nota:** **Criar gráfico** está disponível somente quando **Potência** é selecionada como a suposição de teste **Estimativa** e **Enumeração binomial** não é selecionada.

11. Opcionalmente, clique em **Precisão** para estimar o tamanho da amostra com base em intervalos de confiança, especificando os valores médios do intervalo de confiança. Para obter mais informações, consulte “Análise de energia: precisão” na página 37.

**Nota:** A **precisão** está disponível somente quando o **tamanho da amostra** é selecionado como a suposição de teste Método de **estimativa** e a análise **não direcional (frente e verso)** é selecionada como a **direção do teste**.

### **Análise de potência de binômio de uma amostra: Gráfico**

A caixa de diálogo **Criar gráfico** oferece opções para controlar os gráficos que são produzidos para ilustrar o poder bidimensional e tridimensional dos gráficos. A caixa de diálogo também controla os graus de rotação vertical e horizontal para gráficos tridimensionais.

#### **Gráfico bidimensional**

Fornece opções para controlar a estimação de potência bidimensional versus gráficos. Esta configuração é desativada por padrão.

#### **Estimação de potência versus valor de hipótese nula**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de valor nulo. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

### **Estimação de potência versus valor de hipótese alternativa**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de valores alternativos. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

### **Estimativa de potência versus a diferença entre os valores hipotéticos**

Quando habilitada, esta configuração opcional controla a potência bidimensional por diferença entre o gráfico de valores hipotéticos. Este serviço está desativado por padrão.

### **Estimativa de potência versus o número total de tentativas**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de número total de avaliações. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

### **Intervalo de gráfico do número total de tentativas**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

#### **Limite inferior**

Controla o limite inferior para o gráfico de estimativa de potência bidimensional versus o número total de tentativas. O valor deve ser maior que 0 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para o gráfico de estimativa de potência bidimensional versus o número total de tentativas. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

### **Gráfico tridimensional**

Fornecer opções para controlar a estimação de potência tridimensional versus gráficos. Esta configuração é desativada por padrão.

### **Estimativa de potência versus o número total de tentativas**

Quando selecionada, esta configuração ativa as seguintes opções.

#### **no eixo x e a diferença entre os valores hipotéticos no eixo y**

A configuração opcional controla a potência tridimensional pelo número total de tentativas (eixo x) e diferença entre o gráfico de valores hipotéticos (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

#### **no eixo y e a diferença entre os valores hipotéticos no eixo x**

A configuração opcional controla o poder tridimensional pelo número total de tentativas (eixo y) e diferença entre os valores hipotéticos (eixo x) do gráfico. Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

### **Intervalo de gráfico do número total de tentativas**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

#### **Limite inferior**

Controla o limite inferior para o gráfico de estimativa de potência tridimensional versus o número total de tentativas. O valor deve ser maior que 0 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para o gráfico de estimativa de potência tridimensional versus o número total de tentativas. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

### **Estimação de potência versus valor de hipótese nula**

Quando selecionada, esta configuração ativa as seguintes opções.

### **no eixo x e valor da hipótese alternativa no eixo y**

A configuração opcional controla a potência tridimensional por gráfico nulo (eixo x) e valor alternativo (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

### **no eixo y e valor da hipótese alternativa no eixo x**

A configuração opcional controla a potência tridimensional por gráfico nulo (eixo y) e valor alternativo (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

### **Rotação vertical**

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

### **Rotação horizontal**

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

## **Análise de potência do teste binomial de amostra relacionada**

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

A distribuição binomial é baseada em uma sequência de avaliações de Bernoulli. Ela pode ser usada para modelar experimentos, incluindo um número fixo de tentativas totais, que são consideradas independentes umas das outras. Cada tentativa leva a um resultado dicotômico, com a mesma probabilidade de sucesso.

O binômio de amostra relacionada estima o poder do teste de McNemar para comparar dois parâmetros de proporção com base nos pares pareados de sujeitos amostrados de duas populações binomiais relacionadas.

1. A partir dos menus, escolha:

**Analisar > Análise de potência > Proporções > Teste binomial de amostra relacionada**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.

3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione **Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade”](#) na página 38.

4. Quando a **Potência** for selecionada como o método de **estimativa** de suposição de teste, insira o valor de **Número total de pares** apropriado.

5. Selecione para especificar valores de teste para **Proporções** ou **Contagens**.

- Quando **Proporções** for selecionado, insira valores nos campos **Proporção 1** e **Proporção 2**. Os valores devem estar no intervalo de 0 a 1.
- Quando **Contagens** for selecionado, insira valores nos campos **Contagem 1** e **Contagem 2**. Os valores devem estar entre 0 e o valor especificado para **Número total de pares**.

### Notas de proporções:

- **Proporções** é a única opção disponível quando um valor de **Potência** é especificado.
- Quando os **valores de teste são marginais** não é selecionado:  $0 < \text{Proporção 1} + \text{Proporção 2} \leq 1$
- Quando **Os valores de teste são marginais** estiver selecionado:
  - **Proporção 1 \* Proporção 2** > 0
  - **Proporção 1** < 1
  - **Proporção 2** < 1
  - Os valores para **Proporção 1** e **Proporção 2** não podem ser iguais.

### Notas de contagens:

- As configurações de **contagem** estão disponíveis somente quando **Potência** é selecionada como a configuração **Estimativa** da suposição de teste.
  - Quando os **valores de teste são marginais** não é selecionado:  $0 < \text{Contagem 1} + \text{Contagem 2} \leq \text{Número total de pares}$
  - Quando **Os valores de teste são marginais** estiver selecionado:
    - **Contagem 1 \* Contagem 2** > 0
    - **Contagem 1** < **Número total de pares**
    - **Contagem 2** < **Número total de pares**
6. Opcionalmente, é possível selecionar **Os valores de teste são marginais** para controlar se os valores de proporções ou contagens especificados são ou não marginais. Quando **Os valores de teste são marginais** estiver ativado, deve-se especificar um valor de **Correlação entre pares correspondidos**. O valor deve ser um valor único entre -1 e 1.
7. Selecione um método para estimar a potência.

### Aproximação normal

Ativa a aproximação normal. Essa é a configuração padrão.

### Enumeração binomial

Ativa o método de enumeração binomial. Opcionalmente, use o campo **Limite de tempo** para especificar o número máximo de minutos permitidos para estimar o tamanho da amostra.

Quando o limite de tempo é atingido, a análise é finalizada e uma mensagem de aviso é exibida.

Quando especificado, o valor deve ser um único inteiro positivo para denotar o número de minutos. A configuração padrão é 5 minutos.

8. Selecione se o teste é unilateral ou bilateral.

### Análise não direcional (bilateral)

Quando selecionada, é usado um teste bilateral. Essa é a configuração padrão.

### Análise direcional (unilateral)

Quando selecionada, a potência é calculada para um teste unilateral.

9. Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
10. Opcionalmente, clique em **Plotar** para especificar as configurações de “Análise de potência do teste binomial de amostra relacionada” na página 19 (saída do gráfico, configurações de plotagem bidimensional e configurações de plotagem tridimensional).

**Nota: Criar gráfico** está disponível somente quando **Potência** é selecionada como a suposição de teste **Estimativa** e **Enumeração binomial** não é selecionada.

11. Opcionalmente, clique em **Precisão** para estimar o tamanho da amostra com base em intervalos de confiança, especificando os valores médios do intervalo de confiança. Para obter mais informações, consulte “Análise de energia: precisão” na página 37.

**Nota:** A **precisão** está disponível somente quando o **tamanho da amostra** é selecionado como a suposição de teste Método de **estimativa** e a análise **não direcional (frente e verso)** é selecionada como a **direção do teste**.

## **Análise de potência de binômio de amostra relacionada: Gráfico**

A caixa de diálogo **Gráfico** controla os gráficos que são gerados para ilustrar os gráficos de estimativa de potência bi e tridimensionais. A caixa de diálogo também controla os graus de rotação vertical e horizontal para gráficos tridimensionais.

### **Gráfico bidimensional**

Fornecer opções para controlar a estimação de potência bidimensional versus gráficos. Esta configuração é desativada por padrão.

#### **Estimativa de potência versus o número total de pares**

Quando ativada, esta configuração opcional controla a potência bidimensional pelo número total de pares do gráfico. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

#### **Intervalo de gráfico do número total de pares**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

#### **Limite inferior**

Controla o limite inferior para o gráfico de estimativa de potência bidimensional versus o número total de pares. O valor deve ser maior que 1 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para o gráfico de estimativa de potência bidimensional versus o número total de pares. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 2500.

#### **Estimativa de potência versus diferença de risco**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de diferença de risco. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

#### **Estimativa de potência versus razão de risco**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de razão de risco. Este serviço está desativado por padrão.

#### **Faixa de gráfico de razão de risco**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

#### **Limite inferior**

Controla o limite inferior para o gráfico de estimativa de poder bidimensional versus taxa de risco. O valor não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para o gráfico de estimativa de potência bidimensional versus taxa de risco. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 10.

#### **Estimativa de potência versus razão de chances**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de razão de chances. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

#### **Intervalo de gráfico de razão de chances**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

#### **Limite inferior**

Controla o limite inferior para a estimativa de potência bidimensional versus gráfico de razão de chances. O valor não pode ser maior que o valor de **Limite superior**.

### **Limite superior**

Controla o limite superior para a estimação de potência bidimensional versus gráfico de razão de chances. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 10.

### **Estimativa de potência versus correlação entre pares correspondidos**

Controla a potência estimada  $\Omega$  versus a correlação entre o gráfico de pares combinados. O gráfico é criado apenas quando as proporções marginais ou contagens são especificadas (em vez das proporções discordantes).

### **Gráfico tridimensional**

Fornece opções para controlar a estimação de potência tridimensional versus gráficos. Esta configuração é desativada por padrão.

### **Estimativa de potência versus proporções discordantes**

Quando ativada, esta configuração opcional controla a potência tridimensional por gráfico de proporções discordantes. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

### **Estimativa de potência versus proporções marginais**

Quando ativada, esta configuração opcional controla a potência tridimensional por gráfico de proporções marginais. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

**Nota:** Esta configuração está disponível somente quando **Os valores de teste são marginais** é selecionado.

### **Rotação vertical**

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

### **Rotação horizontal**

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

## **Análise de potência do teste binomial de amostra independente**

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

A distribuição binomial é baseada em uma sequência de avaliações de Bernoulli. Ele pode ser usado para modelar esses experimentos, incluindo um número fixo de tentativas totais que são consideradas independentes umas das outras. Cada tentativa leva a um resultado dicotômico, com a mesma probabilidade de um resultado "bem-sucedido". O teste binomial de amostra independente compara dois parâmetros de proporção independentes.

1. A partir dos menus, escolha:

**Analisar > Análise de potência > Proporções > Teste binomial de amostras independentes**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.

3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione

**Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade”](#) na página 38.

Opcionalmente, especifique um valor de **Proporção do tamanho do grupo**. O valor padrão é 1.

- Quando a **Potência** for selecionada como a configuração **Estimativa** da suposição de teste, insira valores para especificar o **Número total de tentativas para o grupo 1 e o grupo 2**. Os valores devem ser números inteiros maiores que 1.
- Especifique os parâmetros de proporção para os dois grupos. Ambos os valores devem estar no intervalo de 0 a 1.

**Nota:** Os dois valores não podem ser iguais quando um valor de **Potência** é especificado.

- Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
- Selecione o método de teste desejado:

#### **Teste qui-quadrado**

Estima a potência com base no teste qui-quadrado de Pearson. Essa é a configuração padrão.

##### **O desvio padrão é agrupado**

Essa configuração opcional controla se a estimativa do desvio padrão é agrupada ou não agrupada. Essa configuração é ativada por padrão.

##### **Aplicar correção de continuidade**

Esta configuração opcional controla se a correção de continuidade é usada ou não. Este serviço está desativado por padrão.

#### **Teste T**

Estima a potência com base no teste t de Student.

##### **O desvio padrão é agrupado**

Essa configuração opcional controla se a estimativa do desvio padrão é agrupada ou não agrupada. Essa configuração é ativada por padrão.

#### **Teste de razão de verossimilhança**

Estima a potência com base no teste de razão de verossimilhança.

#### **O teste exato de Fisher**

Estima a potência com base no teste exato de Fisher.

##### **Notas:**

- Em alguns casos, o teste exato de Fisher pode levar uma quantidade de tempo maior para concluir.
- Todos os plots são bloqueados quando o teste exato de Fisher é selecionado.

- Selecione um método para estimar a potência.

#### **Aproximação normal**

Ativa a aproximação normal. Essa é a configuração padrão.

#### **Enumeração binomial**

Ativa o método de enumeração binomial. Opcionalmente, use o campo **Limite de tempo** para especificar o número máximo de minutos permitidos para estimar o tamanho da amostra. Quando o limite de tempo é atingido, a análise é finalizada e uma mensagem de aviso é exibida. Quando especificado, o valor deve ser um único inteiro positivo para denotar o número de minutos. A configuração padrão é 5 minutos.

- Selecione se o teste é unilateral ou bilateral.

#### **Análise não direcional (bilateral)**

Quando selecionada, é usado um teste bilateral. Essa é a configuração padrão.

#### **Análise direcional (unilateral)**

Quando selecionada, a potência é calculada para um teste unilateral.

10. Opcionalmente, você pode clicar em **Criar gráfico** para especificar as configurações de “Análise de potência de teste binomial de amostras independentes: Gráfico” na página 24 (saída do gráfico, configurações de plotagem bidimensional e configurações de plotagem tridimensional).

**Nota:** **Criar gráfico** está disponível somente quando **Potência** é selecionada como a suposição de teste **Estimativa** e **Enumeração binomial** não é selecionada.

11. Opcionalmente, clique em **Precisão** para estimar o tamanho da amostra com base em intervalos de confiança, especificando os valores médios do intervalo de confiança. Para obter mais informações, consulte “Análise de energia: precisão” na página 37.

**Nota:** A **precisão** está disponível somente quando o **tamanho da amostra** é selecionado como a suposição de teste Método de **estimativa** e a análise **não direcional (frente e verso)** é selecionada como a **direção do teste**.

### ***Análise de potência de teste binomial de amostras independentes: Gráfico***

A caixa de diálogo **Criar gráfico** fornece opções para controlar os gráficos que são produzidos para ilustrar os gráficos de estimativa de potência bi e tridimensionais. A caixa de diálogo também controla os graus de rotação vertical e horizontal para gráficos tridimensionais.

#### **Gráfico bidimensional**

Fornecer opções para controlar a estimação de potência bidimensional versus gráficos. Esta configuração é desativada por padrão.

#### **Estimativa de poder versus proporção do tamanho do grupo/estimativa de poder versus tamanho do grupo**

Quando habilitada, esta configuração opcional controla o gráfico de proporção de potência bidimensional por tamanho de grupo. Quando selecionada, esta configuração exibe o gráfico. As configurações adicionais não estarão disponíveis quando vários valores de energia forem especificados (**Estimação de energia versus tamanho do grupo**). Este serviço está desativado por padrão.

#### **Faixa de plotagem da proporção do tamanho do grupo**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

#### **Limite inferior**

Controla o limite inferior para o gráfico de estimativa de potência bidimensional versus o número total de pares. O valor deve ser maior que ,01 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para o gráfico de estimativa de potência bidimensional versus o número total de pares. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 100.

#### **Estimativa de potência versus diferença de risco**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de diferença de risco. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

#### **Estimação de potência versus razão de risco**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de razão de risco. Este serviço está desativado por padrão.

#### **Faixa de gráfico de razão de risco**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

#### **Limite inferior**

Controla o limite inferior para o gráfico de estimativa de poder bidimensional versus taxa de risco. O valor não pode ser maior que o valor de **Limite superior**.



**Limite superior**

Controla o limite superior para o gráfico de estimativa de potência bidimensional versus taxa de risco. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 10.

**Estimativa de potência versus razão de chances**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de razão de chances. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

**Intervalo de gráfico de razão de chances**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

**Limite inferior**

Controla o limite inferior para a estimativa de potência bidimensional versus gráfico de razão de chances. O valor não pode ser maior que o valor de **Limite superior**.

**Limite superior**

Controla o limite superior para a estimativa de potência bidimensional versus gráfico de razão de chances. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 10.

**Gráfico tridimensional**

Fornecer opções para controlar a estimativa de potência tridimensional versus gráficos. Esta configuração é desativada por padrão.

**Estimativa de potência versus proporções**

Quando selecionada, esta configuração opcional fornece as seguintes opções de potência por proporção:

**proporção do grupo 1 no eixo x e proporção do grupo 2 no eixo y**

Controla a potência tridimensional pela proporção do gráfico do Grupo 1 (eixo x) e proporção do gráfico do Grupo 2 (eixo y). Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

**proporção do grupo 1 no eixo y e proporção do grupo 2 no eixo x**

Controla o poder tridimensional por proporção do gráfico do Grupo 2 (eixo x) e proporção do gráfico do Grupo 1 (eixo y). Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

**Estimativa de potência versus tamanhos de grupos**

Quando selecionada, esta configuração opcional fornece as seguintes opções de potência por tamanhos de grupo:

**tamanho do grupo 1 no eixo x e tamanho do grupo 2 no eixo y**

Controla o poder tridimensional pelo número de tentativas no gráfico do Grupo 1 (eixo x) e número de tentativas no gráfico do Grupo 2 (eixo y). Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

**tamanho do grupo 1 no eixo y e tamanho do grupo 2 no eixo x**

Controla o poder tridimensional pelo número de tentativas no gráfico do Grupo 2 (eixo x) e número de tentativas no gráfico do Grupo 1 (eixo y). Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

**Intervalo de gráfico do tamanho do grupo 1**

Quando selecionada, as opções de limite inferior e superior para o intervalo de plotagem do grupo 1 estão disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

**Limite inferior**

Controla o limite inferior para a estimativa de poder bidimensional versus gráfico de razão de chances. O valor deve ser maior ou igual a 2 e não pode ser maior que o valor de **Limite superior**.

**Limite superior**

Controla o limite superior para o gráfico de estimativa de poder bidimensional versus razão de chances. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 2500.

**Intervalo de gráfico do tamanho do grupo 2**

Quando selecionada, as opções de limite inferior e superior para o intervalo de plotagem do grupo 2 estão disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

**Limite inferior**

Controla o limite inferior para a estimativa de poder bidimensional versus gráfico de razão de chances. O valor deve ser maior ou igual a 2 e não pode ser maior que o valor de **Limite superior**.

**Limite superior**

Controla o limite superior para o gráfico de estimativa de poder bidimensional versus razão de chances. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 2500.

**Rotação vertical**

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

**Rotação horizontal**

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

## Correlações

Os seguintes recursos estatísticos estão incluídos no IBM SPSS Statistics Base Edition.

### Análise de potência do teste de correlação de Pearson de uma amostra

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

O coeficiente de correlação produto-momento de Pearson mede a força da associação linear entre duas variáveis aleatórias de escala que se supõe seguir uma distribuição normal bivariada. Por convenção, é uma quantidade adimensional e obtida pela padronização da covariância entre duas variáveis contínuas, variando assim entre -1 e 1.

O teste usa o método assintótico de Fisher para estimar o poder da correlação de Pearson de uma amostra.

1. A partir dos menus, escolha:

**Analisar > Análise de potência > Correlações > Momento do Produto Pearson**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.
3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione

**Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade”](#) na página 38.

- Quando **Potência** for selecionada como a configuração **Estimativa** da suposição de teste, insira o **Valor apropriado do tamanho da amostra em pares**. O valor deve ser um único inteiro maior que 3.
- Insira um valor que especifica o valor de hipótese alternativa do parâmetro de correlação no campo **Parâmetro de correlação de Pearson**. O valor deve ser um único numérico entre -1 e 1.

**Nota:** Quando **Potência** é especificada, o valor do **Parâmetro de correlação de Pearson** não pode ser -1 ou 1 e não pode ser igual ao **valor nulo**.

- Opcionalmente, insira um valor que especifica o valor de hipótese nula do parâmetro de correlação a ser testado no campo **Valor nulo**. O valor deve ser um único numérico entre -1 e 1. O valor padrão é 0.

**Nota:** Quando **Potência** é especificada, o **Valor nulo** não pode ser -1 ou 1.

- Opcionalmente, selecione **Usar fórmula de correção de propensão na estimação de potência** para especificar se o ajuste de propensão será envolvido ou ignorado. A configuração é ativada por padrão, o que inclui o termo de ajuste de polarização na estimativa de potência. Quando a configuração não é selecionada, o termo de ajuste de polarização é ignorado.
- Selecione se o teste é unilateral ou bilateral.

#### **Análise não direcional (bilateral)**

Quando selecionada, é usado um teste bilateral. Essa é a configuração padrão.

#### **Análise direcional (unilateral)**

Quando selecionada, a potência é calculada para um teste unilateral.

- Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
- Opcionalmente, é possível clicar em **Criar gráfico** para especificar as configurações de [“Análise de potência de correlação de Pearson de uma amostra: Gráfico”](#) na página 27 (saída do gráfico, configurações de plotagem bidimensional e configurações de plotagem tridimensional).

**Nota:** O **Criar gráfico** está disponível somente quando **Potência** é selecionada como a suposição de teste.

- Opcionalmente, clique em **Precisão** para estimar o tamanho da amostra com base em intervalos de confiança, especificando os valores médios do intervalo de confiança. Para obter mais informações, consulte [“Análise de energia: precisão”](#) na página 37.

**Nota:** A **precisão** está disponível somente quando o **tamanho da amostra** é selecionado como a suposição de teste Método de **estimativa** e a análise **não direcional (frente e verso)** é selecionada como a **direção do teste**.

### ***Análise de potência de correlação de Pearson de uma amostra: Gráfico***

A caixa de diálogo **Criar gráfico** oferece opções para controlar os gráficos que são produzidos para ilustrar a potência bidimensional e tridimensional dos gráficos. A caixa de diálogo também controla os graus de rotação vertical e horizontal para gráficos tridimensionais.

#### **Gráfico bidimensional**

Fornece opções para controlar a estimação de potência bidimensional versus gráficos. As configurações são desativadas por padrão.

#### **Estimação de potência versus valor de hipótese nula**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de valor nulo. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

### **Estimação de potência versus valor de hipótese alternativa**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de valores alternativos. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

### **Estimativa de potência versus a diferença entre os valores hipotéticos**

Quando habilitada, esta configuração opcional controla a potência bidimensional por diferença entre o gráfico de valores hipotéticos. Este serviço está desativado por padrão.

### **Estimativa de potência versus tamanho da amostra (em pares)**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de tamanho da amostra. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

### **Intervalo de gráfico do tamanho da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

#### **Limite inferior**

Controla o limite inferior para a estimação de potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior ou igual a 4 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para a estimação de potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

### **Gráfico tridimensional**

Fornecer opções para controlar a estimação de potência tridimensional versus gráficos. Esta configuração é desativada por padrão.

### **Estimação de potência versus tamanho da amostra**

Quando selecionada, esta configuração ativa as seguintes opções.

#### **no eixo x e a diferença entre os valores hipotéticos no eixo y**

A configuração opcional controla o poder tridimensional por tamanho da amostra (eixo x) e diferença entre o gráfico de valores hipotéticos (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

#### **no eixo y e a diferença entre os valores hipotéticos no eixo x**

A configuração opcional controla a potência tridimensional por tamanho da amostra (eixo y) e diferença entre o gráfico de valores hipotéticos (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

### **Intervalo de gráfico de tamanho da amostra (em pares)**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

#### **Limite inferior**

Controla o limite inferior para a estimativa de potência tridimensional por gráfico de tamanho de amostra. O valor deve ser maior ou igual a 4 e não pode ser maior que o valor de **Limite superior**.

#### **Limite superior**

Controla o limite superior para a estimativa de potência tridimensional por gráfico de tamanho de amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

### **Estimação de potência versus valor de hipótese nula**

Quando selecionada, esta configuração ativa as seguintes opções.

### **no eixo x e valor da hipótese alternativa no eixo y**

A configuração opcional controla a potência tridimensional por gráfico nulo (eixo x) e valor alternativo (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

### **no eixo y e valor da hipótese alternativa no eixo x**

A configuração opcional controla a potência tridimensional por gráfico nulo (eixo y) e valor alternativo (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

### **Rotação vertical**

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

### **Rotação horizontal**

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

## **Análise de potência do teste de correlação de Spearman de uma amostra**

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

O coeficiente de correlação de ordem de classificação de Spearman é uma estatística não paramétrica baseada em classificação para medir a relação monotônica entre duas variáveis que geralmente são censuradas e não distribuídas normalmente. A correlação de ordem de classificação de Spearman é igual à correlação de Pearson entre os valores de classificação das duas variáveis, portanto, também variando entre -1 e 1. Detectar o poder do teste de correlação de postos de Spearman é um tópico importante na análise de dados de séries temporais hidrológicas.

O teste usa o método assintótico de Fisher para estimar o poder da correlação de ordem de classificação de Spearman de uma amostra.

1. A partir dos menus, escolha:

**Analisar > Análise de Potência > Correlações > Ordem de classificação de Spearman**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.
3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione **Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade”](#) na página 38.

4. Quando **Potência** for selecionada como a configuração **Estimativa** da suposição de teste, insira o **Valor apropriado do tamanho da amostra em pares**. O valor deve ser um único inteiro maior que 3.
5. Insira um valor que especifica o valor de hipótese alternativa do parâmetro de correlação no campo **Parâmetro de correlação de Spearman**. O valor deve ser um único numérico entre -1 e 1.

**Nota:** Quando **Potência** é especificado, o valor do **parâmetro de correlação de Spearman** não pode ser -1 ou 1 e não pode ser igual ao **valor nulo**.

6. Opcionalmente, insira um valor que especifica o valor de hipótese nula do parâmetro de correlação a ser testado no campo **Valor nulo**. O valor deve ser um único numérico entre -1 e 1. O valor padrão é 0.

**Nota:** Quando **Potência** é especificada, o **Valor nulo** não pode ser -1 ou 1.

7. Opcionalmente, selecione uma opção que determine como a variância assintótica é estimada para a análise de poder.

#### **Bonett e Wright**

Estima a variância sugerida por Bonett e Wright. Essa é a configuração padrão.

#### **Fieller, Hartley e Pearson**

Estima a variância sugerida por Fieller, Hartley e Pearson.

#### **Caruso e Cliff**

Estima a variância sugerida por Caruso e Cliff.

8. Selecione se o teste é unilateral ou bilateral.

#### **Análise não direcional (bilateral)**

Quando selecionada, é usado um teste bilateral. Essa é a configuração padrão.

#### **Análise direcional (unilateral)**

Quando selecionada, a potência é calculada para um teste unilateral.

9. Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
10. Opcionalmente, clique em **Criar gráfico** para especificar as configurações de “Análise de potência de correlação de Spearman de uma amostra: Gráfico” na página 30 (saída do gráfico, configurações de plotagem bidimensional e configurações de plotagem tridimensional).

**Nota:** O **Criar gráfico** está disponível somente quando **Potência** é selecionada como a suposição de teste.

11. Opcionalmente, clique em **Precisão** para estimar o tamanho da amostra com base em intervalos de confiança, especificando os valores médios do intervalo de confiança. Para obter mais informações, consulte “Análise de energia: precisão” na página 37.

**Nota:** A **precisão** está disponível somente quando o **tamanho da amostra** é selecionado como a suposição de teste Método de **estimativa** e a análise **não direcional (frente e verso)** é selecionada como a **direção do teste**.

### ***Análise de potência de correlação de Spearman de uma amostra: Gráfico***

A caixa de diálogo **Criar gráfico** oferece opções para controlar os gráficos que são produzidos para ilustrar a potência bidimensional e tridimensional dos gráficos. A caixa de diálogo também controla os graus de rotação vertical e horizontal para gráficos tridimensionais.

#### **Gráfico bidimensional**

Fornece opções para controlar a estimação de potência bidimensional versus gráficos. As configurações são desativadas por padrão.

#### **Estimação de potência versus valor de hipótese nula**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de valor nulo. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

#### **Estimação de potência versus valor de hipótese alternativa**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de valores alternativos. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

#### **Estimativa de potência versus a diferença entre os valores hipotéticos**

Quando habilitada, esta configuração opcional controla a potência bidimensional por diferença entre o gráfico de valores hipotéticos. Este serviço está desativado por padrão.

### **Estimativa de potência versus tamanho da amostra (em pares)**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de tamanho da amostra. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

#### **Intervalo de gráfico do tamanho da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

##### **Limite inferior**

Controla o limite inferior para a estimação de potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior ou igual a 4 e não pode ser maior que o valor de **Limite superior**.

##### **Limite superior**

Controla o limite superior para a estimação de potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

### **Gráfico tridimensional**

Fornecer opções para controlar a estimação de potência tridimensional versus gráficos. Esta configuração é desativada por padrão.

#### **Estimação de potência versus tamanho da amostra**

Quando selecionada, esta configuração ativa as seguintes opções.

##### **no eixo x e a diferença entre os valores hipotéticos no eixo y**

A configuração opcional controla o poder tridimensional por tamanho da amostra (eixo x) e diferença entre o gráfico de valores hipotéticos (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

##### **no eixo y e a diferença entre os valores hipotéticos no eixo x**

A configuração opcional controla a potência tridimensional por tamanho da amostra (eixo y) e diferença entre o gráfico de valores hipotéticos (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

#### **Intervalo de gráfico de tamanho da amostra (em pares)**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

##### **Limite inferior**

Controla o limite inferior para a estimativa de potência tridimensional por gráfico de tamanho de amostra. O valor deve ser maior ou igual a 4 e não pode ser maior que o valor de **Limite superior**.

##### **Limite superior**

Controla o limite superior para a estimativa de potência tridimensional por gráfico de tamanho de amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

#### **Estimação de potência versus valor de hipótese nula**

Quando selecionada, esta configuração ativa as seguintes opções.

##### **no eixo x e valor da hipótese alternativa no eixo y**

A configuração opcional controla a potência tridimensional por gráfico nulo (eixo x) e valor alternativo (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

##### **no eixo y e valor da hipótese alternativa no eixo x**

A configuração opcional controla a potência tridimensional por gráfico nulo (eixo y) e valor alternativo (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

### **Rotação vertical**

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

### **Rotação horizontal**

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

## **Análise de potência de teste de correlação de Pearson parcial**

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

A correlação parcial pode ser explicada como a associação entre duas variáveis aleatórias após eliminar o efeito de outra ou várias outras variáveis. É uma medida útil na presença de fatores de confusão. Semelhante ao coeficiente de correlação de Pearson, o coeficiente de correlação parcial também é uma quantidade sem dimensão que varia entre -1 e 1.

O teste usa o método assintótico de Fisher para estimar o poder da correlação de Pearson de uma amostra.

1. A partir dos menus, escolha:

**Analisar > Análise de potência > Correlações > Parcial**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.
3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione **Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade”](#) na página 38.

4. Quando **Potência** é selecionado como a configuração de suposição de teste **Estimativa**, insira o valor do **Tamanho da amostra** apropriado. O valor deve ser um número inteiro único maior que 1.
5. Insira um valor que especifique o **Número das variáveis que se supõe que sejam parcialmente desfeitas**. O valor deve ser um número inteiro único maior ou igual a 0.
6. Insira um valor que especifique o valor da hipótese alternativa do **parâmetro Correlação parcial**. O valor deve ser um único numérico entre -1 e 1.

**Nota:** Quando **Potência** é especificado, o valor de **Parâmetro de correlação parcial** não pode ser -1 ou 1 e não pode ser igual ao **Valor nulo**.

7. Opcionalmente, insira um valor que especifica o valor de hipótese nula do parâmetro de correlação parcial a ser testado no campo **Valor nulo**. O valor deve ser um único numérico entre -1 e 1. O valor padrão é 0.

**Nota:** Quando **Potência** é especificado, o **Valor nulo** não pode ser -1 ou 1.

8. Selecione se o teste é unilateral ou bilateral.

### **Análise não direcional (bilateral)**

Quando selecionada, é usado um teste bilateral. Essa é a configuração padrão.



### **Análise direcional (unilateral)**

Quando selecionada, a potência é calculada para um teste unilateral.

9. Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
10. Opcionalmente, clique em **Plotar** para especificar as configurações de “Análise de potência de correlação de Pearson parcial: Gráfico” na página 33 (saída do gráfico, configurações de plotagem bidimensional e configurações de plotagem tridimensional).

**Nota:** O **Criar gráfico** está disponível somente quando **Potência** é selecionada como a suposição de teste.

11. Opcionalmente, clique em **Precisão** para estimar o tamanho da amostra com base em intervalos de confiança, especificando os valores médios do intervalo de confiança. Para obter mais informações, consulte “Análise de energia: precisão” na página 37.

**Nota:** A **precisão** está disponível somente quando o **tamanho da amostra** é selecionado como a suposição de teste Método de **estimativa** e a análise **não direcional (frente e verso)** é selecionada como a **direção do teste**.

### **Análise de potência de correlação de Pearson parcial: Gráfico**

A caixa de diálogo **Gráfico** oferece opções para controlar os gráficos que são produzidos para ilustrar o poder bidimensional e tridimensional dos gráficos. A caixa de diálogo também controla os graus de rotação vertical e horizontal para gráficos tridimensionais.

#### **Gráfico bidimensional**

Fornece opções para controlar a estimação de potência bidimensional versus gráficos. As configurações são desativadas por padrão.

##### **Estimação de potência versus valor de hipótese nula**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de valor nulo. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

##### **Estimação de potência versus valor de hipótese alternativa**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de valores alternativos. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

##### **Estimativa de potência versus o número de variáveis parcializadas**

Quando habilitada, esta configuração opcional controla o gráfico de potência bidimensional pelo número de variáveis descontinuadas. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

##### **Estimativa de potência versus a diferença entre os valores hipotéticos**

Quando habilitada, esta configuração opcional controla a potência bidimensional por diferença entre o gráfico de valores hipotéticos. Este serviço está desativado por padrão.

##### **Estimação de potência versus tamanho da amostra**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de tamanho da amostra. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

##### **Intervalo de gráfico do tamanho da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

##### **Limite inferior**

Controla o limite inferior para a estimação de potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior ou igual a 4 e não pode ser maior que o valor de **Limite superior**.

### **Limite superior**

Controla o limite superior para a estimação de potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

### **Gráfico tridimensional**

Fornecer opções para controlar a estimação de potência tridimensional versus gráficos. Esta configuração é desativada por padrão.

#### **Estimação de potência versus tamanho da amostra**

Quando selecionada, esta configuração ativa as seguintes opções.

##### **no eixo x e a diferença entre os valores hipotéticos no eixo y**

A configuração opcional controla o poder tridimensional por tamanho da amostra (eixo x) e diferença entre o gráfico de valores hipotéticos (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

##### **no eixo y e a diferença entre os valores hipotéticos no eixo x**

A configuração opcional controla a potência tridimensional por tamanho da amostra (eixo y) e diferença entre o gráfico de valores hipotéticos (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

#### **Intervalo de gráfico de tamanho da amostra (em pares)**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

##### **Limite inferior**

Controla o limite inferior para a estimativa de potência tridimensional por gráfico de tamanho de amostra. O valor deve ser maior ou igual a 4 e não pode ser maior que o valor de **Limite superior**.

##### **Limite superior**

Controla o limite superior para a estimativa de potência tridimensional por gráfico de tamanho de amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

#### **Estimação de potência versus valor de hipótese nula**

Quando selecionada, esta configuração ativa as seguintes opções.

##### **no eixo x e valor da hipótese alternativa no eixo y**

A configuração opcional controla a potência tridimensional por gráfico nulo (eixo x) e valor alternativo (eixo y). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

##### **no eixo y e valor da hipótese alternativa no eixo x**

A configuração opcional controla a potência tridimensional por gráfico nulo (eixo y) e valor alternativo (eixo x). Por padrão, o gráfico é suprimido. Quando especificada, o gráfico é exibido.

#### **Rotação vertical**

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

#### **Rotação horizontal**

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

## Regressão

Os seguintes recursos estatísticos estão incluídos no IBM SPSS Statistics Base Edition.

### Análise de potência de teste de regressão linear univariado

Este recurso requer o IBM SPSS Statistics Base Edition.

A análise de potência desempenha uma função essencial em um plano de estudo, design e condução. O cálculo do poder geralmente é feito antes de quaisquer dados amostrais serem coletados, exceto possivelmente de um pequeno estudo piloto. A estimativa precisa do poder pode dizer aos investigadores quão provável é que uma diferença estatisticamente significativa seja detectada com base em um tamanho de amostra finito sob uma hipótese alternativa verdadeira. Se a potência for muito baixa, há pouca chance de detectar uma diferença significativa, e resultados não significativos são prováveis, mesmo que existam diferenças reais.

A regressão linear univariada é uma abordagem estatística básica e padrão na qual os pesquisadores usam os valores de várias variáveis para explicar ou prever valores de um resultado de escala.

A Análise de potência de teste de regressão linear univariada estima a potência do teste  $F$  de tipo III em vários modelos de regressão linear univariada. Com o tamanho do efeito representado por múltiplas correlações (parciais), são fornecidas abordagens para preditores fixos e aleatórios. Para preditores fixos, a estimação de potência é baseada na distribuição  $F$  não central. Para preditores aleatórios, assume-se que a variável alvo e os preditores seguem conjuntamente uma distribuição normal multivariada. Nesse caso, a estimativa do poder é baseada na distribuição do coeficiente de correlação múltipla amostral.

1. No menu, escolha:

**Analisar > Análise de potência > Regressão > Linear Univariado**

2. Selecione uma configuração de suposição de teste **Estimativa (Tamanho da amostra ou Potência)**.
3. Quando **Tamanho da amostra** é selecionado, insira um **Valor único de potência** para o valor de estimativa de tamanho de amostra (o valor deve ser um valor único entre 0 e 1) ou selecione **Valores de potência da grade** e, em seguida, clique em **Grade** para visualizar tamanhos de amostra projetados para uma gama de valores de potência específicos.

Para obter mais informações, consulte [“Análise de energia: valores de grade”](#) na página 38.

4. Quando **Fazer Estimativa de Potência** for selecionada, insira um **Tamanho de Amostra** apropriado para o valor da estimativa de potência. O valor deve ser um número inteiro único maior ou igual ao número total de preditores de modelo +2 (quando **Incluir o termo de intercepto no modelo** está ativado). Caso contrário, o valor deve ser um único inteiro maior ou igual ao número total de preditores de modelo +1.
5. Especifique o valor do coeficiente de correlação parcial múltipla no campo **Correlação parcial múltipla de população**. O valor deve ser um valor único entre -1 e 1.

**Nota:** Quando um valor de **Potência** é especificado, o valor de **Correlação parcial múltipla de população** não pode ser 0.

As seguintes configurações são ativadas quando **Correlação parcial múltipla de população** é selecionado:

#### Número total de preditores no modelo

Especifique o número dos preditores totais ou dos preditores no modelo completo (não incluindo a interceptação, se aplicável). O valor deve ser um único inteiro maior ou igual a 1.

#### Número de preditores de teste

Especifique o número dos preditores de teste ou dos preditores no modelo aninhado (sem incluir a interceptação, se aplicável). O valor deve ser maior ou igual a 1, mas não maior que o valor de **Número total de preditores no modelo**.

6. Especifique **Valores de R quadrado para** coeficientes de correlação múltipla para **Modelo completo** e **Modelo agrupado**. Os valores devem ser um único valor entre 0 e 1.

**Nota:** Quando um valor de **Potência** é especificado, o valor de **Modelo completo** deve ser maior que o valor de **Modelo agrupado**.

As configurações a seguir são ativadas quando **Valores de R quadrado para** é selecionado:

**Número total de preditores - Modelo completo**

Especifique o número de preditores totais para o modelo completo (não incluindo a interceptação, se aplicável). O valor deve ser um único inteiro maior ou igual a 1.

**Número total de preditores - Modelo aninhado**

Especifique o número de preditores totais para o modelo aninhado (sem incluir a interceptação, se aplicável). O valor deve ser maior ou igual a 1, mas menor que o número **Total de preditores - valor do modelo completo**.

7. Opcionalmente, especifique o nível de significância da taxa de erro Tipo I para o teste no campo **Nível de significância**. O valor deve ser um valor duplo único entre 0 e 1. O valor padrão é 0,05.
8. Opcionalmente, é possível selecionar a configuração **Incluir o termo de intercepto no modelo**. Essa configuração é ativada por padrão. Quando não selecionado, o termo de interceptação é excluído da análise de potência.
9. Opcionalmente, é possível selecionar se os preditores de modelo são **Fixos** ou **Aleatórios**. **Fixo** é a configuração padrão.
10. Opcionalmente, é possível clicar em **Gráfico** para especificar configurações do “Análise de potência de regressão linear univariada: Gráfico” na página 36 (saída de gráfico, configurações de gráfico bidimensional e configurações de gráfico tridimensional).

**Nota:** O **Criar gráfico** está disponível somente quando **Potência** é selecionada como a suposição de teste.

## ***Análise de potência de regressão linear univariada: Gráfico***

É possível controlar os gráficos que são emitidos para ilustrar o poder bidimensional e tridimensional por gráficos. Também é possível controlar a exibição de dicas de ferramentas e os graus de rotação vertical/horizontal para gráficos tridimensionais.

### **Gráfico bidimensional**

Fornece opções para controlar a estimação de potência bidimensional versus gráficos. As configurações são desativadas por padrão.

#### **Estimativa de poder versus a correlação parcial múltipla**

Quando ativada, esta configuração opcional controla o poder bidimensional pelo gráfico de coeficiente de correlação parcial múltipla. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

#### **Estimativa de potência versus tamanho da amostra**

Quando ativada, esta configuração opcional controla a potência bidimensional por gráfico de tamanho da amostra. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

#### **Intervalo de plotagem do tamanho da amostra**

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

##### **Limite inferior**

Controla o limite inferior para a estimação de potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior ou igual a 4 e não pode ser maior que o valor de **Limite superior**.

##### **Limite superior**

Controla o limite superior para a estimação de potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

## Gráfico tridimensional

Fornecer opções para controlar a estimativa de potência tridimensional versus gráficos, as configurações de rotação vertical/horizontal e o intervalo de tamanho de amostra do gráfico especificado pelo usuário. Esta configuração está desativada por padrão.

### Estimativa de potência versus tamanho da amostra

Quando ativada, esta configuração opcional controla o poder tridimensional por gráficos de tamanho de amostra. Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

### no eixo x e a correlação parcial múltipla no eixo y

Esta configuração opcional controla o poder tridimensional por tamanho de amostra (eixo x) e gráfico de coeficiente de correlação parcial múltipla (eixo y). Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

### no eixo y e a correlação parcial múltipla no eixo x

Esta configuração opcional controla o poder tridimensional por tamanho de amostra (eixo y) e gráfico de coeficiente de correlação parcial múltipla (eixo x). Este serviço está desativado por padrão. Quando selecionada, esta configuração exibe o gráfico.

### Intervalo de plotagem do tamanho da amostra

Quando selecionada, as opções de limite inferior e superior ficam disponíveis. Quando nenhum valor de número inteiro for especificado para os campos **Limite inferior** ou **Limite superior**, o intervalo de gráfico padrão será usado.

#### Limite inferior

Controla o limite inferior para a estimação de potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior ou igual a 4 e não pode ser maior que o valor de **Limite superior**.

#### Limite superior

Controla o limite superior para a estimação de potência bidimensional por gráfico de tamanho da amostra. O valor deve ser maior que o valor de **Limite inferior** e não pode ser maior que 5000.

### Rotação vertical

A configuração opcional define os graus de rotação vertical (no sentido horário a partir da esquerda) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico verticalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 10.

### Rotação horizontal

A configuração opcional define os graus de rotação horizontal (no sentido horário a partir da frente) para o gráfico tridimensional. É possível usar o mouse para girar o gráfico horizontalmente. A configuração entra em vigor quando a plotagem tridimensional é solicitada. O valor deve ser um único valor inteiro menor ou igual a 359. O valor padrão é 325.

## Análise de energia: precisão

A caixa de diálogo **Precisão** está disponível para todos os procedimentos de Análise de Energia (exceto para Regressão Linear Univariada) e fornece opções para estimar o tamanho da amostra com base nas meias larguras do intervalo de confiança especificado. A caixa de diálogo está disponível quando **Valor de potência única** ou **Valores de potência de grade** são especificados para a estimativa do **tamanho da amostra** e quando a configuração de **Direção do teste** é configurada como **Não direcional (análise de dois lados)**.

**Nota:** A configuração de **Direção do teste** não se aplica ao ANOVA de uma via.

quando as opções **Estimar o tamanho da amostra** e **valores de potência da grade** são selecionadas (clique no controle **Grade** para exibir a caixa de diálogo).

### Tipo(s) de intervalo(s) de confiança

Selecione os tipos de intervalos de confiança apropriados e se a correção de continuidade é aplicada.

**Nota: Tipo(s) de intervalo de confiança** estão disponíveis apenas para os procedimentos de proporção de Análise de Potência (Proporção de Uma Amostra, Teste Binomial de Amostras Relacionadas e Teste Binomial de Amostras Independentes).

### **Especificar a metade da largura do intervalo de confiança**

Estima o tamanho da amostra com base no valor de meia largura do intervalo de confiança. Insira um valor no intervalo de 0 a 1. Para o teste Binomial de Uma Amostra, o valor deve estar no intervalo de 0 a 0,5.

**Nota:** Valores duplicados são ignorados.

- Clique em **Incluir** para incluir o valor de meia largura especificado na lista.
- Realce um valor de meia largura existente e clique em **Mudar** para atualizar o valor.
- Destaque um valor de meia largura existente e clique em **Remover** para remover o valor da lista.

## **Análise de energia: valores de grade**

O diálogo **Valores de grade** fornece opções para especificar um intervalo de valores **POWER** com o propósito de visualizar tamanhos da amostra projetados em um formato de grade para cada valor de intervalo **POWER** especificado.

O diálogo **Valores de grade** está disponível para cada procedimento de Análise de energia quando as opções **Estimar tamanho da amostra** e **Valores de grade de energia** são selecionadas (clique no controle **Grade** para exibir o diálogo).

### **Especificar potência única**

Quando selecionada, pelo menos um valor é necessário para executar a análise. São permitidos diversos valores e cada valor deve pertencer a  $[0, 1]$ . É possível especificar diversos valores, com cada valor separado por um ou mais espaços em branco. Use os controles **Incluir**, **Mudar** e **Remover** para trabalhar com os valores na lista de valores de energia.

Todos os valores devem ser exclusivos (não são permitidos valores duplicados).

### **Especificar intervalo de potência**

Quando selecionado, um intervalo de valores de energia pode ser especificado de um valor **Inicial** (value1) para um valor **Final** (value2) com o incremento de **Por** (value3). Se especificado, somente um conjunto válido de  $[\text{value1 } T0 \text{ value2 } BY \text{ value3}]$  é permitido. Ele deve satisfazer esse  $0 \leq \text{value1} \leq \text{value2} \leq 1$ . Nos casos em que  $\text{value1} = \text{value2}$ , isso é equivalente a especificar um único value1, independentemente de value3.

**Nota:** As opções **Especificar energia única** e **Especificar intervalo de energia** são independentes; é possível selecionar uma única opção ou ambas as opções.

## **Meta-análise**

---

A meta-análise é a análise dos dados obtidos a partir de uma coleção de estudos que respondem a questões de pesquisa semelhantes. Esses estudos são conhecidos como estudos primários. A meta-análise usa métodos estatísticos para produzir uma estimativa geral de um efeito, explorar a heterogeneidade entre os estudos e investigar o impacto do viés de publicação ou, mais geralmente, efeitos de pequenos estudos nos resultados finais.

O IBM SPSS Statistics suporta tamanhos de efeito padrão e tamanhos de efeito genéricos (pré-calculados) para dados binários (como a razão de chances de log) e para dados contínuos (como g de Hedges). As informações de meta-análise (como os tamanhos de efeito específicos do estudo e seus erros padrão correspondentes e o modelo e método de meta-análise) são especificadas durante a etapa de declaração de meta-análise. As informações são usadas automaticamente por todas as meta-análises subsequentes.

Modelos de meta-análise de efeitos aleatórios, efeitos comuns e efeitos fixos são suportados. Dependendo do modelo de meta-análise escolhido, os vários métodos de estimativa (por exemplo, variância inversa e Mantel–Haenszel) estão disponíveis para os modelos de efeito comum e efeitos fixos.

Vários estimadores diferentes também estão disponíveis para o parâmetro de variância entre estudos para o modelo de efeitos aleatórios.

O IBM SPSS Statistics suporta os procedimentos de meta-análise a seguir:

- [“Metanálise Contínua” na página 39](#)
- [“Tamanho do Efeito Contínuo de Metanálise” na página 47](#)
- [“Binário de Metanálise” na página 55](#)
- [“Tamanho do Efeito Binário de Metanálise” na página 64](#)
- [“Regressão de Metanálise” na página 73](#)

## Metanálise Contínua

O procedimento de Metanálise Contínua executa a metanálise com resultados contínuos em dados brutos que são fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### Exemplo

Vários estudos de pesquisa foram realizados no passado para investigar um medicamento da moda, porém questionável, para ajudar no tratamento de diabetes tipo II. O medicamento oral foi alegado ser capaz de reduzir o nível de glicose no sangue após as refeições. Os dados foram coletados em diferentes locais de pesquisa de 1979 a 1986.

Um investigador principal gostaria de fazer inferência estatística sobre o efeito do medicamento oral. Devido ao fato de que os dados foram gerados a partir de diferentes estudos, ela propôs a ideia de sintetizar os resultados entre os estudos para alcançar uma compreensão geral do efeito e identificar as fontes subjacentes de variação nos resultados.

### Estatísticas

Nível de confiança, método iterativo, redução de etapa pela metade, tolerância de convergência, médias da amostra, variância da amostra, desvio padrão, tamanho do efeito estimado,  $d$  do Cohen,  $g$  do Hedges, Delta do Glass, diferença média, análise acumulativa, método de estimação, trim-and-fill, teste baseado em regressão, modelo de efeitos aleatórios, modelo de efeitos fixos, estimador de máxima verossimilhança restrita, estimador de Bayes empírico, estimador de Hedges, estimador de Hunter-Schmidt, estimador de DerSimonian-Laird, estimador de Sidik-Jonkman, ajuste de erro padrão de Knapp-Hartung, ajuste de erro padrão de Knapp-Hartung truncado, coeficientes, teste baseado em regressão da EGGER, intercepto, modelo multiplicativo, parâmetro de dispersão multiplicativo, estimador quadrático, teste de homogeneidade, medidas de heterogeneidade, intervalo de predição, erro padrão estimado, valor  $p$  estimado, tamanho do efeito geral acumulativo, peso de estudo estimado.

## Obtendo uma análise de Metanálise contínua

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Dados Brutos...**

2. Sob a seção **Grupo de tratamento**, selecione uma variável de **Tamanho do estudo** para representar o tamanho da amostra para o grupo de tratamento. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).
3. Selecione uma variável de **Média** para representar as médias da amostra para o grupo de tratamento. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).
4. Selecione **Desvio padrão** para especificar o desvio padrão de amostra ou **Varição** para especificar a variância da amostra e, em seguida, selecione uma variável para representar o desvio/a variação padrão para o grupo de tratamento.
5. Sob a seção **Grupo de controle**, selecione uma variável de **Tamanho do estudo** para representar o tamanho da amostra para o grupo de controle. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).
6. Selecione uma variável de **Média** para representar as médias da amostra para o grupo de controle. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).

7. Selecione **Desvio padrão** para especificar o desvio padrão de amostra ou **Varição** para especificar a variância da amostra e, em seguida, selecione uma variável para representar o desvio/a variação padrão para o grupo de controle.
8. Opcionalmente, selecione as variáveis **ID do Estudo** e/ou **Rótulo do Estudo**. A variável **ID do Estudo** selecionada não pode ser igual à variável **Rótulo do Estudo** selecionada.
9. Opcionalmente, selecione a configuração **Tamanho do Efeito**.

#### **d de Cohen**

A configuração padrão estima o *d* de Cohen. Quando o **Erro padrão ajustado** é selecionado, a configuração estima o *d* de Cohen e sua variância usando uma fórmula alternativa dividida por  $2(N_{\text{treatment}} + N_{\text{control}} - 2)$ .

#### **G de Hedges**

Estima o *g* de Hedges. Quando o **Erro padrão ajustado** é selecionado, a configuração estima o *g* de Hedge e sua variância usando uma fórmula alternativa dividida por  $2(N_{\text{treatment}} + N_{\text{control}} - 3,94)$ .

#### **Delta de Glass**

Estima o Delta do Glass com base no grupo de controle. Quando **Padronizado baseado em grupo de tratamento** é selecionado, o Delta do Glass é padronizado com base no desvio padrão do grupo de tratamento.

#### **Diferença média não padronizada**

Estima a diferença média assumindo que os dois desvios padrão de preenchimento são iguais. Quando **Variâncias de grupos desiguais** são selecionadas, a diferença média é estimada assumindo que os dois desvios padrão de preenchimento não são iguais.

10. Opcionalmente, selecione uma configuração **Modelo**. Quando as configurações **Aparar e Preencher** são ativadas, elas também controlam o modelo usado, agrupando a análise aparar e preencher. Quando as configurações **Viés** são ativadas, elas também controlam o modelo usado pelo teste baseado em regressão.

#### **Efeitos aleatórios**

A configuração padrão desenvolve o modelo de efeitos aleatórios.

#### **Efeitos fixos**

Desenvolve o modelo de efeitos fixos.

11. Opcionalmente, é possível:
  - Clique em **Crítérios...** para especificar os critérios gerais.
  - Clique em **Análise** para especificar o subgrupo e a análise acumulativa.
  - Clique em **Inferência** para especificar os métodos de estimação.
  - Clique em **Contraste** para controlar o teste de contraste.
  - Clique em **Viés** para acessar o viés de publicação, realizando o teste baseado em regressão da EGGER.
  - Clique em **Aparar e Preencher** para implementar a análise aparar e preencher do viés da publicação.
  - Clique em **Imprimir** para controlar as saídas de tabela.
  - Clique em **Salvar** para salvar as estatísticas estimadas para o conjunto de dados ativo.
  - Clique em **Gráfico** para especificar quais gráficos incluir na saída.
12. Clique em **OK**.

## **Metanálise contínua: critérios**

O diálogo **Crítérios** fornece configurações para especificar critérios para metanálise com resultados contínuos em dados brutos que são fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.



### **Intervalo de confiança**

A configuração opcional especifica o nível de confiança. O valor deve ser um valor numérico entre 0 e 100. A configuração padrão é 95.

### **Escopo de dados omissos**

As configurações opcionais controlam como o procedimento manipula dados ausentes.

#### **Excluir análise de casos por análise**

A configuração padrão inclui todos os casos com dados suficientes sobre as variáveis usadas em cada análise específica.

#### **Excluir casos de listwise**

Inclui todos os casos com dados suficientes sobre todas as variáveis usadas em todas as análises especificadas pelo procedimento.

### **Valores ausentes do usuário**

As configurações opcionais controlam como os valores ausentes do usuário são tratados.

#### **Exclude**

A configuração padrão trata os valores ausentes do usuário como válidos.

#### **Include**

Ignora as designações de valor omissos do usuário e trata os valores omissos do usuário como válidos.

### **Iterações**

#### **Iterações máximas**

A configuração opcional especifica o número máximo de iterações nos métodos iterativos. A configuração está disponível quando métodos iterativos são usados. O valor deve ser um único número inteiro positivo. O valor-padrão é 100. Um valor 0 significa que não são realizadas iterações.

#### **Redução máxima da etapa pela metade**

A configuração opcional especifica a redução máxima pela metade da etapa nos métodos iterativos. A configuração está disponível quando métodos iterativos são usados. O valor deve ser um único número inteiro positivo. O valor padrão é 5 segundos. Um valor de 0 significa que a redução da etapa pela metade não é aplicada.

#### **Convergência**

A configuração opcional especifica a tolerância de convergência. O valor deve ser um valor positivo único. A configuração está disponível quando métodos iterativos são usados. O valor padrão é  $1E-6$ .

## **Definindo critérios de Metanálise contínua**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Dados Brutos...**

2. No diálogo **Metanálise contínua**, clique em **Critérios**.

3. Selecione e defina as configurações apropriadas de critérios.

4. Clique **Continuar**.

## **Metanálise contínua: análise**

O diálogo **Análise** fornece configurações para especificar análise de subgrupo e acumulativa para metanálise com resultados contínuos em dados brutos que são fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### **Análise de subgrupo**

Selecione uma variável que evoque a análise de subgrupo. A variável não pode ser a mesma que a variável definida para **Análise acumulativa**.

### **Análise acumulativa**

Selecione uma variável que evoca a análise acumulativa e sobre a qual a meta-análise acumulativa é conduzida. A variável não pode ser a mesma que a variável definida para **Análise de subgrupo**. Quando **Crescente** é selecionado, a análise acumulativa é baseada na variável especificada em ordem crescente. Quando **Decrescente** é selecionado, a análise acumulativa é baseada na variável especificada em ordem decrescente.

### **Estatísticas acumulativas**

Fornece opções para salvar o tamanho do efeito geral acumulativo estimado. As configurações estão disponíveis apenas quando uma variável de **Análise acumulativa** é selecionada.

#### **Tamanho do efeito acumulativo**

Salva o tamanho do efeito geral acumulativo estimado.

#### **Erro padrão**

Salva o erro padrão estimado do tamanho do efeito geral acumulativo.

#### **Límite inferior de intervalo de confiança**

Salva o limite inferior do intervalo de confiança estimado do tamanho do efeito geral acumulativo.

#### **Límite superior de intervalo de confiança**

Salva o limite superior do intervalo de confiança estimado do tamanho do efeito geral acumulativo.

#### **Valor *p***

Salva o valor *p* estimado do tamanho do efeito geral acumulativo.

### **Destino**

Fornece opções para especificar um conjunto de dados ou arquivo de dados salvo. Ao selecionar **Conjunto de dados**, é possível especificar um novo nome de conjunto de dados ou manter o nome padrão. Ao selecionar **Arquivo de dados**, clique em **Navegar...** para selecionar o nome e o local de um arquivo salvo.

## **Definindo configurações de análise de Metanálise contínua**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Dados Brutos...**

2. No diálogo **Metanálise contínua**, clique em **Análise**.

3. Selecione e defina as configurações apropriadas de análise.

4. Clique **Continuar**.

## **Metanálise contínua: inferência**

O diálogo **Inferência** fornece configurações para especificar os métodos de estimação para a metanálise com resultados contínuos em dados brutos fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

**Nota:** O diálogo **Inferência** está disponível somente quando um modelo **Efeitos Aleatórios** é selecionado.

### **Estimador**

Fornece configurações para especificar o estimador.

#### **Notas:**

- Quando as configurações de Aparar e Preencher são especificadas, também controla o estimador usado pelo agrupamento na análise de aparar e preencher.
- Quando as configurações de Viés são especificadas, também controla o estimador usado pelo teste baseado em regressão.

#### **Máxima verossimilhança restrita (REML)**

A configuração padrão aplica o método iterativo e calcula o estimador de máxima verossimilhança restrito.

**Máxima verossimilhança (ML)**

Aplica o método iterativo e calcula o estimador de máxima verossimilhança.

**Bayes empírico**

Aplica o método iterativo e calcula o estimador Bayes empírico.

**Hedges**

Aplica o método não iterativo e calcula o estimador de Hedges.

**Hunter-Schmidt**

Aplica o método não iterativo e calcula o estimador de Hunter-Schmidt.

**DerSimonian-Laird**

Aplica o método não iterativo e calcula o estimador DerSimonian-Laird.

**Sidik-Jonkman**

Aplica o método não iterativo e calcula o estimador de Sidik-Jonkman.

**Ajuste de erro padrão**

Fornecer configurações que controlam se o ajuste de erro padrão de Knapp-Hartung deve ser aplicado.

**Notas:**

- Quando as configurações de corte e preenchimento são especificadas, também controla o ajuste de erro padrão que é usado pelo agrupamento na análise de corte e preenchimento.
- Quando as configurações de Bias são especificadas, também controla o ajuste de erro padrão usado pelo teste baseado em regressão.

**Nenhum ajuste**

A configuração padrão não aplica o ajustamento.

**Aplicar o ajuste de Knapp-Hartung**

Aplica o método de ajustamento Knapp-Hartung.

**Aplicar o ajuste truncado de Knapp-Hartung**

Aplica o método de ajuste de Knapp-Hartung e trunca o valor se for menor que 1 ao estimar a matriz de variância-covariância.

**Definindo configurações de Inferência contínua de metanálise**

1. No menu, escolha:

**Analisar > Metanálise > Resultados Contínuos > Dados Brutos...**

2. No diálogo **Metanálise contínua**, clique em **Inferência**.

3. Selecione e defina as configurações apropriadas de inferência.

4. Clique **Continuar**.

**Metanálise contínua: contraste**

O diálogo **Contraste** fornece configurações para controlar o teste de contraste para metanálise, com resultados contínuos em dados brutos que são fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

**Variáveis**

A lista exibe todas as variáveis de conjunto de dados disponíveis. Selecione variáveis na lista e mova-as para a lista de **Contrastes**.

**Contrastes**

A lista identifica os coeficientes armazenados como variáveis no conjunto de dados ativo. Várias variáveis são permitidas. Variáveis de sequência de caracteres não são suportadas.

**Valores de coeficientes de entrada do usuário**

Fornecer configurações para especificar coeficientes de um contraste fornecido pelo usuário. Apenas valores numéricos são permitidos. Para formular um teste de contraste válido, o número de valores especificados deve corresponder ao dos estudos válidos.

## Definindo configurações de contraste de Metanálise contínua

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Dados Brutos...**

2. No diálogo **Metanálise contínua**, clique em **Contraste**.

3. Selecione e defina as configurações apropriadas de contraste.

4. Clique **Continuar**.

## Metanálise contínua: viés

A caixa de diálogo **Propensão** fornece configurações para ativar a propensão de publicação realizando o teste baseado em regressão de Eggers para metanálise, com resultados contínuos de dados brutos fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### Teste baseado em regressão de Egger

A seleção dessa configuração ativa o viés de publicação realizando o teste baseado em regressão de Eggers.

### Variáveis

A lista fornece todas as variáveis de conjunto de dados disponíveis.

### Covariáveis

Variáveis selecionadas na lista de **Variáveis** são tratadas como covariáveis. Várias covariáveis são permitidas.

### Fatores

Variáveis selecionadas na lista de **Variáveis** são tratadas como fatores. Vários fatores são permitidos.

### Incluir interceptação na regressão

Controla o termo de intercepto no teste baseado em regressão.

### Incluir parâmetro de dispersão em modelo de efeitos fixos

Controla a configuração de modelo multiplicativo e apresenta o parâmetro de dispersão multiplicativa para a análise. A configuração está disponível apenas quando um modelo de efeitos fixos é selecionado.

### Estimar estatísticas com base na distribuição t

Controla a distribuição usada nos testes baseados em regressão. A configuração é ativada por padrão e estima as estatísticas com base em *t*-distribution. Se a configuração não for selecionada, as estatísticas serão estimadas com base na distribuição normal.

## Definindo configurações de viés de Metanálise contínua

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Dados Brutos...**

2. No diálogo **Metanálise contínua**, clique em **Viés**.

3. Selecione e defina as configurações apropriadas de viés.

4. Clique **Continuar**.

## Metanálise contínua: trim-and-fill

O diálogo **Trim-and-Fill** fornece configurações para implementar a análise de trim-and-fill de viés de publicação para metanálise com resultados contínuos em dados brutos que são fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### Estimar o número de estudos omissos

Controla a análise de trim-and-fill de viés de publicação. A seleção dessa configuração habilita as outras configurações de diálogo.

### Estudos colaterais para imputação

Fornecer opções para especificar o lado do gráfico de funil no qual os estudos omissos são imputados.

**Determinado pelas inclinações do teste de Egger**

A configuração padrão determina o lado com base na inclinação estimada do teste de Egger.

**Esquerda**

Imputa o lado esquerdo do gráfico de funil.

**Direita**

Imputa o lado direito do gráfico de funil.

**Método**

Especifica o método para estimar o número de estudos ausentes.

**Linear**

A configuração padrão calcula o estimador linear.

**Executar o**

Calcula o estimador de execução.

**Quadrático**

Calcula o estimador quadrático.

**Processo de iteração**

Fornecer configurações para especificar o estimador de iteração e o ajuste de erro padrão.

**Modelo de efeitos fixos**

Quando selecionado, um modelo de efeitos fixos é empregado e as opções de estimativa de iteração e ajuste de erro padrão não estão disponíveis.

**Modelo de efeitos aleatórios**

Quando selecionado, um modelo de efeitos aleatórios é empregado e as seguintes configurações estão disponíveis.

**Estimador**

Fornecer configurações para especificar o estimador de iteração

**Máxima verossimilhança restrita (REML)**

A configuração padrão aplica o método iterativo e calcula o estimador de máxima verossimilhança restrito.

**Máxima verossimilhança (ML)**

Aplica o método iterativo e calcula o estimador de máxima verossimilhança.

**Bayes empírico**

Aplica o método iterativo e calcula o estimador Bayes empírico.

**Hedges**

Aplica o método não iterativo e calcula o estimador de Hedges

**Hunter-Schmidt**

Aplica o método não iterativo e calcula o estimador de Hunter-Schmidt.

**DerSimonian-Laird**

Aplica o método não iterativo e calcula o estimador DerSimonian-Laird.

**Sidik-Jonkman**

Aplica o método não iterativo e calcula o estimador de Sidik-Jonkman.

**Ajuste de erro padrão**

Fornecer configurações para controlar se o ajuste de erro padrão de Knapp-Hartung deve ser aplicado às iterações do algoritmo de trim-and-fill.

**Nenhum ajuste**

A configuração padrão não aplica o ajustamento.

**Aplicar o ajuste de Knapp-Hartung**

Aplica o método de ajustamento Knapp-Hartung.

**Aplicar o ajuste truncado de Knapp-Hartung**

Aplica o método de ajuste de Knapp-Hartung e trunca o valor se for menor que 1 ao estimar a matriz de variância-covariância.

## Definindo configurações de trim-and-fill de Metanálise contínua

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Dados Brutos...**

2. No diálogo **Metanálise contínua**, clique em **Trim-and-Fill**.

3. Selecione e defina as configurações apropriadas de aparar e preencher.

4. Clique **Continuar**.

## Metanálise contínua: imprimir

O diálogo **Imprimir** fornece configurações para controlar saídas de tabela para metanálise com resultados contínuos em dados brutos que são fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### Homogeneidade/Heterogeneidade

Fornecer configurações para controlar os testes de homogeneidade e heterogeneidade.

#### Teste de homogeneidade

Quando selecionado, o teste de homogeneidade correspondente é fornecido na saída.

#### Medidas de heterogeneidade

Quando selecionadas, as medidas de heterogeneidade são fornecidas na saída.

### Tamanhos do efeito

Fornecer as configurações de tamanho do efeito a seguir.

#### Estudos individuais

Controla a exibição de estudos individuais. Quando selecionado, a saída correspondente é fornecida.

#### Tamanhos de efeito acumulativo

Controla a exibição de análise acumulativa. Quando selecionado, a saída correspondente é fornecida. A configuração está disponível apenas quando uma variável de **Análise acumulativa** é selecionada no diálogo **Análise**.

#### Intervalo de predição sob o modelo de efeitos aleatórios

Controla a exibição do intervalo de predição. Quando selecionado, a saída correspondente é fornecida. A configuração está disponível somente quando um modelo de efeitos aleatórios é especificado.

## Definindo configurações de impressão de Metanálise contínua

1. No menu, escolha:

**Analisar > Metanálise > Resultados Contínuos > Dados Brutos...**

2. No diálogo **Metanálise contínua**, clique em **Imprimir**.

3. Selecione e defina as configurações apropriadas de impressão.

4. Clique **Continuar**.

## Metanálise contínua: salvar

O diálogo **Salvar** fornece configurações para salvar as estatísticas estimadas no conjunto de dados ativo para metanálise, com resultados contínuos em dados brutos fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### Estudos individuais

Fornecer opções para salvar o tamanho do efeito estimado.

#### Tamanho do efeito individual

Salva o tamanho do efeito estimado.

**Erro padrão**

Salva o erro padrão estimado do tamanho do efeito.

**Limite inferior de intervalo de confiança**

Salva o limite inferior de intervalo de confiança estimado do tamanho do efeito.

**Limite superior de intervalo de confiança**

Salva o limite superior de intervalo de confiança estimado do tamanho do efeito.

**Valor  $p$** 

Salva o valor  $p$  estimado do tamanho do efeito.

**Ponderação de estudo**

Salva o peso do estudo estimado.

**Porcentagem de ponderação do estudo**

Salva o peso do estudo normalizado como uma porcentagem.

**Definindo configurações de salvamento de Metanálise contínua**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Dados Brutos...**

2. No diálogo **Metanálise contínua**, clique em **Salvar**.

3. Selecione e defina as configurações apropriadas de salvamento das estatísticas estimadas.

4. Clique **Continuar**.

**Metanálise Contínua: Gráfico**

O diálogo **Gráfico** fornece configurações para os tipos de gráfico a seguir:

- “Gráfico de floresta” na página 77
- “Gráfico de floresta acumulativo” na página 78
- “Gráfico de bolhas” na página 79
- “Gráfico de funil” na página 80
- “Gráfico de Galbraith” na página 81

**Tamanho do Efeito Contínuo de Metanálise**

O procedimento de tamanho de efeito contínuo de meta-análise executa meta-análise com resultados contínuos quando os dados de tamanho de efeito pré-calculados são fornecidos no conjunto de dados ativo.

**Exemplo**

Vários estudos de pesquisa foram realizados na história para investigar um medicamento da moda, mas discutível, para ajudar a tratar o diabetes tipo II. O medicamento oral foi alegado ser capaz de reduzir o nível de glicose no sangue após as refeições. Os dados foram coletados em diferentes locais de pesquisa de 1979 a 1986.

Um investigador principal gostaria de fazer inferência estatística sobre o efeito do medicamento oral. Devido ao fato de que os dados foram gerados a partir de diferentes estudos, ela propôs a ideia de sintetizar os resultados entre os estudos para alcançar uma compreensão geral do efeito e identificar as fontes subjacentes de variação nos resultados.

**Estatísticas**

Nível de confiança, método iterativo, redução de etapa pela metade, tolerância de convergência, médias da amostra, variância da amostra, desvio padrão, tamanho do efeito estimado,  $d$  do Cohen,  $g$  do Hedges, Delta do Glass, diferença média, análise acumulativa, método de estimação, trim-and-fill, teste baseado em regressão, modelo de efeitos aleatórios, modelo de efeitos fixos, estimador de máxima verossimilhança restrita, estimador de Bayes empírico, estimador de Hedges, estimador de Hunter-Schmidt, estimador de DerSimonian-Laird, estimador de Sidik-Jonkman, ajuste de erro padrão de Knapp-Hartung, ajuste de erro padrão de Knapp-Hartung truncado, coeficientes, teste baseado

em regressão da EGGER, intercepto, modelo multiplicativo, parâmetro de dispersão multiplicativo, estimador quadrático, teste de homogeneidade, medidas de heterogeneidade, intervalo de predição, erro padrão estimado, valor  $p$  estimado, tamanho do efeito geral acumulativo, peso de estudo estimado,

## Obtendo uma análise de Tamanho do efeito contínuo de metanálise

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Tamanho do Efeito Pré-calculado**

2. Selecione uma variável de **Tamanho do efeito** que denote o tamanho do efeito. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).
3. Selecione **Desvio padrão** para especificar o desvio padrão de tamanho do efeito ou **Variação** para especificar a variação de tamanho do efeito e, em seguida, selecione uma variável para representar o desvio/a variação padrão para o tamanho do efeito.
4. Opcionalmente, selecione as variáveis **ID do Estudo** e/ou **Rótulo do Estudo**. A variável **ID do Estudo** selecionada não pode ser igual à variável **Rótulo do Estudo** selecionada.
5. Opcionalmente, selecione uma configuração **Modelo**. Quando as configurações **Aparar e Preencher** são ativadas, elas também controlam o modelo usado, agrupando a análise aparar e preencher. Quando as configurações **Viés** são ativadas, elas também controlam o modelo usado pelo teste baseado em regressão.

### Efeitos aleatórios

A configuração padrão desenvolve o modelo de efeitos aleatórios.

### Efeitos fixos

Constrói o modelo de efeitos fixos.

6. Opcionalmente, é possível:

- Clique em **Critérios...** para especificar os critérios gerais.
- Clique em **Análise** para especificar o subgrupo e a análise acumulativa.
- Clique em **Inferência** para especificar os métodos de estimação.
- Clique em **Contraste** para controlar o teste de contraste.
- Clique em **Viés** para acessar o viés de publicação, realizando o teste baseado em regressão da EGGER.
- Clique em **Aparar e Preencher** para implementar a análise aparar e preencher do viés da publicação.
- Clique em **Imprimir** para controlar as saídas de tabela.
- Clique em **Salvar** para salvar as estatísticas estimadas para o conjunto de dados ativo.
- Clique em **Gráfico** para especificar quais gráficos incluir na saída.

7. Clique em **OK**.

## Tamanho do efeito contínuo de metanálise: critérios

O diálogo **Critérios** fornece configurações para especificar critérios para metanálise com resultados contínuos quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### Intervalo de confiança

A configuração opcional especifica o nível de confiança. O valor deve ser um valor numérico entre 0 e 100. A configuração padrão é 95.

### Escopo de dados omissos

As configurações opcionais controlam como o procedimento manipula dados ausentes.



**Excluir análise de casos por análise**

A configuração padrão inclui todos os casos com dados suficientes sobre as variáveis usadas em cada análise específica.

**Excluir casos de listwise**

Inclui todos os casos com dados suficientes sobre todas as variáveis usadas em todas as análises especificadas pelo procedimento.

**Valores ausentes do usuário**

As configurações opcionais controlam como os valores ausentes do usuário são tratados.

**Exclui**

A configuração padrão trata os valores ausentes do usuário como válidos.

**Inclui**

Ignora as designações de valor omissos do usuário e trata os valores omissos do usuário como válidos.

**Iterações****Iterações máximas**

A configuração opcional especifica o número máximo de iterações nos métodos iterativos. A configuração está disponível quando métodos iterativos são usados. O valor deve ser um único número inteiro positivo. O valor-padrão é 100. Um valor 0 significa que não são realizadas iterações.

**Redução máxima da etapa pela metade**

A configuração opcional especifica a redução máxima pela metade da etapa nos métodos iterativos. A configuração está disponível quando métodos iterativos são usados. O valor deve ser um único número inteiro positivo. O valor padrão é 5 segundos. Um valor de 0 significa que a redução da etapa pela metade não é aplicada.

**Convergência**

A configuração opcional especifica a tolerância de convergência. O valor deve ser um valor positivo único. A configuração está disponível quando métodos iterativos são usados. O valor padrão é 1E-6.

**Definindo critérios de Tamanho do efeito contínuo de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Tamanho do Efeito Pré-calculado**

2. No diálogo **Tamanho do efeito contínuo de metanálise**, clique em **Critérios**.

3. Selecione e defina as configurações apropriadas de critérios.

4. Clique **Continuar**.

**Tamanho do efeito contínuo de metanálise: análise**

O diálogo **Análise** fornece configurações para especificar análise de subgrupo e acumulativa para metanálise com resultados contínuos quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

**Análise de subgrupo**

Selecione uma variável que evoque a análise de subgrupo. A variável não pode ser a mesma que a variável definida para **Análise acumulativa**.

**Análise acumulativa**

Selecione uma variável que evoca a análise acumulativa e sobre a qual a meta-análise acumulativa é conduzida. A variável não pode ser a mesma que a variável definida para **Análise de subgrupo**. Quando **Crescente** é selecionado, a análise acumulativa é baseada na variável especificada em ordem crescente. Quando **Decrescente** é selecionado, a análise acumulativa é baseada na variável especificada em ordem decrescente.

### **Estatísticas acumulativas**

Fornecer opções para salvar o tamanho do efeito geral acumulativo estimado. As configurações estão disponíveis apenas quando uma variável de **Análise acumulativa** é selecionada.

#### **Tamanho do efeito acumulativo**

Salva o tamanho do efeito geral acumulativo estimado.

#### **Erro padrão**

Salva o erro padrão estimado do tamanho do efeito geral acumulativo.

#### **Limite inferior de intervalo de confiança**

Salva o limite inferior do intervalo de confiança estimado do tamanho do efeito geral acumulativo.

#### **Limite superior de intervalo de confiança**

Salva o limite superior do intervalo de confiança estimado do tamanho do efeito geral acumulativo.

#### **Valor p**

Salva o valor  $p$  estimado do tamanho do efeito geral acumulativo.

### **Destino**

Fornecer opções para especificar um conjunto de dados ou arquivo de dados salvo. Ao selecionar **Conjunto de dados**, é possível especificar um novo nome de conjunto de dados ou manter o nome padrão. Ao selecionar **Arquivo de dados**, clique em **Navegar...** para selecionar o nome e o local de um arquivo salvo.

## **Definindo configurações de análise de Tamanho do efeito contínuo de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Tamanho do Efeito Pré-calculado**

2. No diálogo **Tamanho do efeito contínuo de metanálise**, clique em **Análise**.

3. Selecione e defina as configurações apropriadas de análise.

4. Clique **Continuar**.

## **Tamanho do efeito contínuo de metanálise: inferência**

O diálogo **Inferência** fornece configurações para especificar os métodos de estimação para metanálise com resultados contínuos quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

**Nota:** O diálogo **Inferência** está disponível somente quando um modelo **Efeitos Aleatórios** é selecionado.

### **Estimador**

Fornecer configurações para especificar o estimador.

#### **Notas:**

- Quando as configurações de Aparar e Preencher são especificadas, também controla o estimador usado pelo agrupamento na análise de aparar e preencher.
- Quando as configurações de Viés são especificadas, também controla o estimador usado pelo teste baseado em regressão.

#### **Máxima verossimilhança restrita (REML)**

A configuração padrão aplica o método iterativo e calcula o estimador de máxima verossimilhança restrito.

#### **Máxima verossimilhança (ML)**

Aplica o método iterativo e calcula o estimador de máxima verossimilhança.

#### **Bayes empírico**

Aplica o método iterativo e calcula o estimador Bayes empírico.

**Hedges**

Aplica o método não iterativo e calcula o estimador de Hedges.

**Hunter-Schmidt**

Aplica o método não iterativo e calcula o estimador de Hunter-Schmidt.

**DerSimonian-Laird**

Aplica o método não iterativo e calcula o estimador DerSimonian-Laird.

**Sidik-Jonkman**

Aplica o método não iterativo e calcula o estimador de Sidik-Jonkman.

**Ajuste de erro padrão**

Fornece configurações que controlam se o ajuste de erro padrão de Knapp-Hartung deve ser aplicado.

**Notas:**

- Quando as configurações de corte e preenchimento são especificadas, também controla o ajuste de erro padrão que é usado pelo agrupamento na análise de corte e preenchimento.
- Quando as configurações de Bias são especificadas, também controla o ajuste de erro padrão usado pelo teste baseado em regressão.

**Nenhum ajuste**

A configuração padrão não aplica o ajustamento.

**Aplicar o ajuste de Knapp-Hartung**

Aplica o método de ajustamento Knapp-Hartung.

**Aplicar o ajuste truncado de Knapp-Hartung**

Aplica o método de ajuste de Knapp-Hartung e trunca o valor se for menor que 1 ao estimar a matriz de variância-covariância.

**Definindo configurações de inferência do Tamanho do efeito contínuo de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Tamanho do Efeito Pré-calculado**

2. No diálogo **Tamanho do efeito contínuo de metanálise**, clique em **Inferência**.
3. Selecione e defina as configurações apropriadas de inferência.
4. Clique **Continuar**.

**Tamanho do efeito contínuo de metanálise: contraste**

O diálogo **Contraste** fornece configurações para controlar o teste de contraste para metanálise com resultados contínuos quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

**Variáveis**

A lista exibe todas as variáveis de conjunto de dados disponíveis. Selecione variáveis na lista e mova-as para a lista de **Contrastes**.

**Contrastes**

A lista identifica os coeficientes armazenados como variáveis no conjunto de dados ativo. Várias variáveis são permitidas. Variáveis de sequência de caracteres não são suportadas.

**Valores de coeficientes de entrada do usuário**

Fornece configurações para especificar coeficientes de um contraste fornecido pelo usuário. Apenas valores numéricos são permitidos. Para formular um teste de contraste válido, o número de valores especificados deve corresponder ao dos estudos válidos.

**Definindo configurações de contraste de Tamanho do efeito contínuo de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Tamanho do Efeito Pré-calculado**

2. No diálogo **Tamanho do efeito contínuo de metanálise**, clique em **Contraste**.
3. Selecione e defina as configurações apropriadas de contraste.
4. Clique **Continuar**.

## Tamanho do efeito contínuo de metanálise: viés

A caixa de diálogo **Propensão** fornece configurações para ativar a propensão de publicação realizando o teste baseado em regressão de Eggers para metanálise, com resultados contínuos quando os dados pré-calculados de tamanho do efeito são fornecidos no conjunto de dados ativo.

### Teste baseado em regressão de Egger

A seleção dessa configuração ativa o viés de publicação realizando o teste baseado em regressão de Eggers.

### Variáveis

A lista fornece todas as variáveis de conjunto de dados disponíveis.

### Covariáveis

Variáveis selecionadas na lista de **Variáveis** são tratadas como covariáveis. Várias covariáveis são permitidas.

### Fatores

Variáveis selecionadas na lista de **Variáveis** são tratadas como fatores. Vários fatores são permitidos.

### Incluir interceptação na regressão

Controla o termo de intercepto no teste baseado em regressão.

### Incluir parâmetro de dispersão em modelo de efeitos fixos

Controla a configuração de modelo multiplicativo e apresenta o parâmetro de dispersão multiplicativa para a análise. A configuração está disponível apenas quando um modelo de efeitos fixos é selecionado.

### Estimar estatísticas com base na distribuição t

Controla a distribuição usada nos testes baseados em regressão. A configuração é ativada por padrão e estima as estatísticas com base em  $t$ -distribution. Se a configuração não for selecionada, as estatísticas serão estimadas com base na distribuição normal.

## Definindo configurações de viés de Tamanho do efeito contínuo de metanálise

1. A partir dos menus, escolha:  
**Analisar > Metanálise > Resultados Contínuos > Tamanho do Efeito Pré-calculado**
2. No diálogo **Tamanho do efeito contínuo de metanálise**, clique em **Viés**.
3. Selecione e defina as configurações apropriadas de viés.
4. Clique **Continuar**.

## Tamanho do efeito contínuo de metanálise: Trim-and-Fill

O diálogo **Trim-and-Fill** fornece configurações para implementar a análise de trim-and-fill de viés de publicação para metanálise com resultados contínuos quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### Estimar o número de estudos omissos

Controla a análise de trim-and-fill de viés de publicação. A seleção dessa configuração habilita as outras configurações de diálogo.

### Estudos colaterais para imputação

Fornece opções para especificar o lado do gráfico de funil no qual os estudos omissos são imputados.

#### Determinado pelas inclinações do teste de Egger

A configuração padrão determina o lado com base na inclinação estimada do teste de Egger.

#### Esquerda

Imputa o lado esquerdo do gráfico de funil.

**Direita**

Imputa o lado direito do gráfico de funil.

**Método**

Especifica o método para estimar o número de estudos ausentes.

**Linear**

A configuração padrão calcula o estimador linear.

**Executar o**

Calcula o estimador de execução.

**Quadrático**

Calcula o estimador quadrático.

**Processo de iteração**

Fornecer configurações para especificar o estimador de iteração e o ajuste de erro padrão.

**Modelo de efeitos fixos**

Quando selecionado, um modelo de efeitos fixos é empregado e as opções de estimativa de iteração e ajuste de erro padrão não estão disponíveis.

**Modelo de efeitos aleatórios**

Quando selecionado, um modelo de efeitos aleatórios é empregado e as seguintes configurações estão disponíveis.

**Estimador**

Fornecer configurações para especificar o estimador de iteração

**Máxima verossimilhança restrita (REML)**

A configuração padrão aplica o método iterativo e calcula o estimador de máxima verossimilhança restrito.

**Máxima verossimilhança (ML)**

Aplica o método iterativo e calcula o estimador de máxima verossimilhança.

**Bayes empírico**

Aplica o método iterativo e calcula o estimador Bayes empírico.

**Hedges**

Aplica o método não iterativo e calcula o estimador de Hedges

**Hunter-Schmidt**

Aplica o método não iterativo e calcula o estimador de Hunter-Schmidt.

**DerSimonian-Laird**

Aplica o método não iterativo e calcula o estimador DerSimonian-Laird.

**Sidik-Jonkman**

Aplica o método não iterativo e calcula o estimador de Sidik-Jonkman.

**Ajuste de erro padrão**

Fornecer configurações para controlar se o ajuste de erro padrão de Knapp-Hartung deve ser aplicado às iterações do algoritmo de trim-and-fill.

**Nenhum ajuste**

A configuração padrão não aplica o ajustamento.

**Aplicar o ajuste de Knapp-Hartung**

Aplica o método de ajustamento Knapp-Hartung.

**Aplicar o ajuste truncado de Knapp-Hartung**

Aplica o método de ajuste de Knapp-Hartung e trunca o valor se for menor que 1 ao estimar a matriz de variância-covariância.

**Definindo configurações de trim-and-fill de Tamanho do efeito contínuo de metanálise**

1. A partir dos menus, escolha:

## **Analisar > Metanálise > Resultados Contínuos > Tamanho do Efeito Pré-calculado**

2. No diálogo **Tamanho do efeito contínuo de metanálise**, clique em **Trim-and-Fill**.
3. Selecione e defina as configurações apropriadas de aparar e preencher.
4. Clique **Continuar**.

## **Tamanho do efeito contínuo de metanálise: imprimir**

O diálogo **Imprimir** fornece configurações para controlar saídas de tabela para metanálise com resultados contínuos quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### **Homogeneidade/Heterogeneidade**

Fornece configurações para controlar os testes de homogeneidade e heterogeneidade.

#### **Teste de homogeneidade**

Quando selecionado, o teste de homogeneidade correspondente é fornecido na saída.

#### **Medidas de heterogeneidade**

Quando selecionadas, as medidas de heterogeneidade são fornecidas na saída.

### **Tamanhos do efeito**

Fornece as configurações de tamanho do efeito a seguir.

#### **Estudos individuais**

Controla a exibição de estudos individuais. Quando selecionado, a saída correspondente é fornecida.

#### **Tamanhos de efeito acumulativo**

Controla a exibição de análise acumulativa. Quando selecionado, a saída correspondente é fornecida. A configuração está disponível apenas quando uma variável de **Análise acumulativa** é selecionada no diálogo **Análise**.

#### **Intervalo de predição sob o modelo de efeitos aleatórios**

Controla a exibição do intervalo de predição. Quando selecionado, a saída correspondente é fornecida. A configuração está disponível somente quando um modelo de efeitos aleatórios é especificado.

## **Definindo configurações de impressão de Tamanho do efeito contínuo de metanálise**

1. A partir dos menus, escolha:

### **Analisar > Metanálise > Resultados Contínuos > Tamanho do Efeito Pré-calculado**

2. No diálogo **Tamanho do efeito contínuo de metanálise**, clique em **Imprimir**.
3. Selecione e defina as configurações apropriadas de impressão.
4. Clique **Continuar**.

## **Tamanho do efeito contínuo de metanálise: salvar**

O diálogo **Salvar** fornece configurações para salvar as estatísticas estimadas para o conjunto de dados ativo para a metanálise, com resultados contínuos quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### **Estudos individuais**

Fornece opções para salvar o tamanho do efeito estimado.

#### **Erro padrão**

Salva o erro padrão estimado do tamanho do efeito. A configuração está disponível apenas quando uma variável de **Tamanho do efeito** não é especificada.

#### **Limite inferior de intervalo de confiança**

Salva o limite inferior de intervalo de confiança estimado do tamanho do efeito.

**Limite superior de intervalo de confiança**

Salva o limite superior de intervalo de confiança estimado do tamanho do efeito.

**Valor  $p$** 

Salva o valor  $p$  estimado do tamanho do efeito.

**Ponderação de estudo**

Salva o peso do estudo estimado.

**Porcentagem de ponderação do estudo**

Salva o peso do estudo normalizado como uma porcentagem.

**Definindo configurações de salvamento do Tamanho do efeito contínuo de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos > Tamanho do Efeito Pré-calculado**

2. No diálogo **Tamanho do efeito contínuo de metanálise**, clique em **Salvar**.

3. Selecione e defina as configurações apropriadas de salvamento das estatísticas estimadas.

4. Clique **Continuar**.

**Tamanho do efeito contínuo de metanálise: criar gráfico**

O diálogo **Gráfico** fornece configurações para os tipos de gráfico a seguir:

- “Gráfico de floresta” na página 77
- “Gráfico de floresta acumulativo” na página 78
- “Gráfico de bolhas” na página 79
- “Gráfico de funil” na página 80
- “Gráfico de Galbraith” na página 81

**Binário de Metanálise**

O procedimento Binário de Metanálise realiza meta-análise com resultados binários em dados brutos que são fornecidos no conjunto de dados ativo para a estimativa do tamanho do efeito.

**Exemplo**

Vários estudos de pesquisa foram realizados na história para investigar um medicamento da moda, mas discutível, para ajudar a tratar o diabetes tipo II. O medicamento oral foi alegado ser capaz de reduzir o nível de glicose no sangue após as refeições. Os dados foram coletados em diferentes locais de pesquisa de 1979 a 1986.

Um investigador principal gostaria de fazer inferência estatística sobre o efeito do medicamento oral. Devido ao fato de que os dados foram gerados a partir de diferentes estudos, ela propôs a ideia de sintetizar os resultados entre os estudos para alcançar uma compreensão geral do efeito e identificar as fontes subjacentes de variação nos resultados.

**Estatísticas**

Intervalo de confiança, Razão de Chances de Log, Razão de Chances de Log de Peto, Razão de Risco de Log, diferença de risco, efeitos aleatórios, efeitos fixos, variância inversa, Mantel-Haenszel, iterações, fracionamento de passos, convergência, estatísticas acumulativas, tamanho do efeito acumulativo, máxima verossimilhança restrita, REML, máxima verossimilhança, ML, Bayes Empírico, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, Teste de Egger, Teste de Harbord, Teste de Peters, interceptação na regressão, parâmetro de dispersão, homogeneidade, heterogeneidade, estatística exponencial, erro padrão, valor  $p$ , ponderação de estudo.

**Obtendo uma análise de Binário de metanálise**

1. A partir dos menus, escolha:

### **Analisar > Metanálise > Resultados Binários > Dados Brutos...**

2. Sob a seção **Grupo de tratamento**, selecione uma variável de **Sucesso** para representar as contagens de “sucesso” para o grupo de tratamento. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).
3. Selecione uma variável de **Falha** para representar as contagens de "falha" para o grupo de tratamento. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).
4. Sob a seção **Grupo de controle**, selecione uma variável de **Sucesso** para representar as contagens de “sucesso” para o grupo de controle. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).
5. Selecione uma variável de **Falha** para representar as contagens de "falha" para o grupo de controle. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).
6. Opcionalmente, selecione as variáveis **ID do Estudo** e/ou **Rótulo do Estudo**. A variável **ID do Estudo** selecionada não pode ser igual à variável **Rótulo do Estudo** selecionada.
7. Opcionalmente, selecione a configuração **Tamanho do Efeito**. As opções disponíveis são **Razão de chances de log**, **Razão de chances de log de Peto**, **Razão de risco de log** e **Diferença de risco**.
8. Opcionalmente, selecione uma configuração **Modelo**. Quando as configurações **Aparar e Preencher** são ativadas, elas também controlam o modelo usado, agrupando a análise aparar e preencher. Quando as configurações **Viés** são ativadas, elas também controlam o modelo usado pelo teste baseado em regressão.

#### **Efeitos aleatórios**

A configuração padrão desenvolve o modelo de efeitos aleatórios.

#### **Efeitos fixos**

Constrói o modelo de efeitos fixos. **Variância inversa** estima o peso da variância inversa. **Mantel-Haenszel** estima o peso de Mantel-Haenszel.

9. Opcionalmente, é possível:
  - Clique em **Crítérios...** para especificar os critérios gerais.
  - Clique em **Análise** para especificar o subgrupo e a análise acumulativa.
  - Clique em **Inferência** para especificar os métodos de estimação.
  - Clique em **Contraste** para controlar o teste de contraste.
  - Clique em **Viés** para acessar o viés de publicação, realizando o teste baseado em regressão da EGGER.
  - Clique em **Aparar e Preencher** para implementar a análise aparar e preencher do viés da publicação.
  - Clique em **Imprimir** para controlar as saídas de tabela.
  - Clique em **Salvar** para salvar as estatísticas estimadas para o conjunto de dados ativo.
  - Clique em **Gráfico** para especificar quais gráficos incluir na saída.
10. Clique em **OK**.

## **Binário de Metanálise: critérios**

O diálogo **Crítérios** fornece configurações para especificar critérios para metanálise com resultados binários em dados brutos que são fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

#### **Intervalo de confiança**

A configuração opcional especifica o nível de confiança. O valor deve ser um valor numérico entre 0 e 100. A configuração padrão é 95.

#### **Escopo de dados omissos**

As configurações opcionais controlam como o procedimento manipula dados ausentes.



**Excluir análise de casos por análise**

A configuração padrão inclui todos os casos com dados suficientes sobre as variáveis usadas em cada análise específica.

**Excluir casos de listwise**

Inclui todos os casos com dados suficientes sobre todas as variáveis usadas em todas as análises especificadas pelo procedimento.

**Valores ausentes do usuário**

As configurações opcionais controlam como os valores ausentes do usuário são tratados.

**Exclui**

A configuração padrão trata os valores ausentes do usuário como válidos.

**Inclui**

Ignora as designações de valor omissos do usuário e trata os valores omissos do usuário como válidos.

**Iterações****Iterações máximas**

A configuração opcional especifica o número máximo de iterações nos métodos iterativos. A configuração está disponível quando métodos iterativos são usados. O valor deve ser um único número inteiro positivo. O valor-padrão é 100. Um valor 0 significa que não são realizadas iterações.

**Redução máxima da etapa pela metade**

A configuração opcional especifica a redução máxima pela metade da etapa nos métodos iterativos. A configuração está disponível quando métodos iterativos são usados. O valor deve ser um único número inteiro positivo. O valor padrão é 5 segundos. Um valor de 0 significa que a redução da etapa pela metade não é aplicada.

**Convergência**

A configuração opcional especifica a tolerância de convergência. O valor deve ser um valor positivo único. A configuração está disponível quando métodos iterativos são usados. O valor padrão é 1E-6.

**Ajustar contagens zero**

As configurações opcionais controlam como ajustar os dados de contagem zero, se aplicável, na estimativa do tamanho do efeito.

**Ajuste cada valor apenas dos estudos contendo pelo menos um zero**

A configuração padrão ajusta apenas dados de contagem zero.

**Ajustar cada valor de todos os estudos se pelo menos um estudo contiver um valor zero**

Ajusta todas as frequências quando pelo menos uma contagem zero existe em um estudo.

**Ajuste cada valor de todos os estudos**

Ajusta todos os dados, independentemente da existência de contagens zero.

**Não faça ajustes e mantenha zero(s)**

Não faz ajustes de dados.

**Valor incluído**

A configuração opcional especifica o valor incluído nos dados de contagem zero. O valor padrão é 0.5. O valor especificado deve ser um único valor numérico maior que 0 e menor ou igual a 1.

**Definindo critérios de Binário de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Dados Brutos...**

2. No diálogo **Binário de metanálise**, clique em **Crítérios**.

3. Selecione e defina as configurações apropriadas de critérios.

4. Clique **Continuar**.

## Binário de metanálise: análise

O diálogo **Análise** fornece configurações para especificar a análise de subgrupo e acumulativa para a metanálise, com resultados binários em dados brutos fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### Análise de subgrupo

Selecione uma variável que evoque a análise de subgrupo. A variável não pode ser a mesma que a variável definida para **Análise acumulativa**.

### Análise acumulativa

Selecione uma variável que evoca a análise acumulativa e sobre a qual a meta-análise acumulativa é conduzida. A variável não pode ser a mesma que a variável definida para **Análise de subgrupo**. Quando **Crescente** é selecionado, a análise acumulativa é baseada na variável especificada em ordem crescente. Quando **Decrescente** é selecionado, a análise acumulativa é baseada na variável especificada em ordem decrescente.

### Estatísticas acumulativas

Fornece opções para salvar o tamanho do efeito geral acumulativo estimado. As configurações estão disponíveis apenas quando uma variável de **Análise acumulativa** é selecionada.

#### Tamanho do efeito acumulativo

Salva o tamanho do efeito geral acumulativo estimado.

#### Tamanho do efeito acumulativo (forma exponenciada)

Salva o tamanho do efeito geral acumulativo estimado na forma exponencial.

#### Erro padrão

Salva o erro padrão estimado do tamanho do efeito geral acumulativo.

#### Limite inferior de intervalo de confiança

Salva o limite inferior do intervalo de confiança estimado do tamanho do efeito geral acumulativo.

#### Limite superior de intervalo de confiança

Salva o limite superior do intervalo de confiança estimado do tamanho do efeito geral acumulativo.

#### Limite inferior do intervalo de confiança (forma exponencial)

Salva o limite inferior do intervalo de confiança estimado do tamanho do efeito geral acumulativo na forma exponencial.

#### Limite superior do intervalo de confiança (forma exponencial)

Salva o limite superior do intervalo de confiança estimado do tamanho do efeito geral acumulativo na forma exponencial.

#### Valor $p$

Salva o valor  $p$  estimado do tamanho do efeito geral acumulativo.

### Destino

Fornece opções para especificar um conjunto de dados ou arquivo de dados salvo. Ao selecionar **Conjunto de dados**, é possível especificar um novo nome de conjunto de dados ou manter o nome padrão. Ao selecionar **Arquivo de dados**, clique em **Navegar...** para selecionar o nome e o local de um arquivo salvo.

## Definindo configurações de análise de Binário de metanálise

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Dados Brutos...**

2. No diálogo **Binário de metanálise**, clique em **Análise**.

3. Selecione e defina as configurações apropriadas de análise.

4. Clique **Continuar**.

## Binário de metanálise: inferência

O diálogo **Inferência** fornece configurações para especificar os métodos de estimação para a metanálise com resultados binários em dados brutos fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

**Nota:** O diálogo **Inferência** está disponível somente quando um modelo **Efeitos Aleatórios** é selecionado.

### Estimador

Fornecer configurações para especificar o estimador.

#### Notas:

- Quando as configurações de Aparar e Preencher são especificadas, também controla o estimador usado pelo agrupamento na análise de aparar e preencher.
- Quando as configurações de Viés são especificadas, também controla o estimador usado pelo teste baseado em regressão.

#### Máxima verossimilhança restrita (REML)

A configuração padrão aplica o método iterativo e calcula o estimador de máxima verossimilhança restrito.

#### Máxima verossimilhança (ML)

Aplica o método iterativo e calcula o estimador de máxima verossimilhança.

#### Bayes empírico

Aplica o método iterativo e calcula o estimador Bayes empírico.

#### Hedges

Aplica o método não iterativo e calcula o estimador de Hedges.

#### Hunter-Schmidt

Aplica o método não iterativo e calcula o estimador de Hunter-Schmidt.

#### DerSimonian-Laird

Aplica o método não iterativo e calcula o estimador DerSimonian-Laird.

#### Sidik-Jonkman

Aplica o método não iterativo e calcula o estimador de Sidik-Jonkman.

### Ajuste de erro padrão

Fornecer configurações que controlam se o ajuste de erro padrão de Knapp-Hartung deve ser aplicado.

#### Notas:

- Quando as configurações de corte e preenchimento são especificadas, também controla o ajuste de erro padrão que é usado pelo agrupamento na análise de corte e preenchimento.
- Quando as configurações de Bias são especificadas, também controla o ajuste de erro padrão usado pelo teste baseado em regressão.

#### Nenhum ajuste

A configuração padrão não aplica o ajustamento.

#### Aplicar o ajuste de Knapp-Hartung

Aplica o método de ajustamento Knapp-Hartung.

#### Aplicar o ajuste truncado de Knapp-Hartung

Aplica o método de ajuste de Knapp-Hartung e trunca o valor se for menor que 1 ao estimar a matriz de variância-covariância.

## Definindo configurações de inferência de Binário de metanálise

1. No menu, escolha:

**Analisar > Metanálise > Resultados Binários > Dados Brutos...**

2. No diálogo **Binário de metanálise**, clique em **Inferência**.

3. Selecione e defina as configurações apropriadas de inferência.
4. Clique **Continuar**.

## Binário de Metanálise: contraste

O diálogo **Contraste** fornece configurações para controlar o teste de contraste para metanálise, com resultados binários em dados brutos fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### Variáveis

A lista exibe todas as variáveis de conjunto de dados disponíveis. Selecione variáveis na lista e mova-as para a lista de **Contrastes**.

### Contrastes

A lista identifica os coeficientes armazenados como variáveis no conjunto de dados ativo. Várias variáveis são permitidas. Variáveis de sequência de caracteres não são suportadas.

### Valores de coeficientes de entrada do usuário

Fornecer configurações para especificar coeficientes de um contraste fornecido pelo usuário. Apenas valores numéricos são permitidos. Para formular um teste de contraste válido, o número de valores especificados deve corresponder ao dos estudos válidos.

### Exibir estatísticas exponenciadas

Controla a inclusão das estatísticas de forma exponencial. Quando a configuração é selecionada, a saída inclui as estatísticas de forma exponencial, incluindo o tamanho do efeito exponencial e os limites do intervalo de confiança. A configuração está disponível quando o **Tamanho do Efeito** é especificado como **Razão de Chances de Logs**, **Razão de Chances de Logs de Peto** ou **Razão de Risco de Log**.

## Definindo configurações de contraste de Binário de metanálise

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Dados Brutos...**

2. No diálogo **Binário de metanálise**, clique em **Contraste**.
3. Selecione e defina as configurações apropriadas de contraste.
4. Clique **Continuar**.

## Binário de metanálise: Viés

O diálogo **Viés** fornece configurações para ativar o viés de publicação realizando testes baseados em regressão para metanálise com resultados binários em dados brutos que são fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### Testes baseados em regressão

Fornecer opções para especificar os testes baseados em regressão. Mais de um teste pode ser selecionado.

#### Teste de Egger

Quando selecionado, realiza o teste da Egger.

#### Teste de Harbord

Quando selecionado, realiza o teste do Harbord. O teste está disponível quando o **Tamanho do efeito** é especificado como **Razão de chances de log** ou **Razão de risco de log**.

#### Teste de Peters

Quando selecionado, realiza o teste de Peters. O teste está disponível quando o **Tamanho do efeito** é especificado como **Razão de chances de logs**.

### Variáveis

A lista fornece todas as variáveis de conjunto de dados disponíveis.

### **Covariáveis**

Variáveis selecionadas na lista de **Variáveis** são tratadas como covariáveis. Várias covariáveis são permitidas.

### **Fatores**

Variáveis selecionadas na lista de **Variáveis** são tratadas como fatores. Vários fatores são permitidos.

### **Incluir interceptação na regressão**

Controla o termo de intercepto no teste baseado em regressão.

### **Incluir parâmetro de dispersão em modelo de efeitos fixos**

Controla a configuração de modelo multiplicativo e apresenta o parâmetro de dispersão multiplicativa para a análise. A configuração está disponível apenas quando um modelo de efeitos fixos é selecionado.

### **Estimar estatísticas com base na distribuição t**

Controla a distribuição usada nos testes baseados em regressão. A configuração é ativada por padrão e estima as estatísticas com base em *t*-distribution. Se a configuração não for selecionada, as estatísticas serão estimadas com base na distribuição normal.

## **Definindo configurações de viés de Binário de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Dados Brutos...**

2. No diálogo **Binário de metanálise**, clique em **Viés**.

3. Selecione e defina as configurações apropriadas de viés.

4. Clique **Continuar**.

## **Binário de metanálise: Trim-and-Fill**

O diálogo **Trim-and-Fill** fornece configurações para a implementação da análise de trim-and-fill de viés de publicação para a metanálise, com resultados binários em dados brutos fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### **Estimar o número de estudos omissos**

Controla a análise de trim-and-fill de viés de publicação. A seleção dessa configuração habilita as outras configurações de diálogo.

### **Estudos colaterais para imputação**

Fornecer opções para especificar o lado do gráfico de funil no qual os estudos omissos são imputados.

#### **Determinado pelas inclinações do teste de Egger**

A configuração padrão determina o lado com base na inclinação estimada do teste de Egger.

#### **Esquerda**

Imputa o lado esquerdo do gráfico de funil.

#### **Direita**

Imputa o lado direito do gráfico de funil.

### **Método**

Especifica o método para estimar o número de estudos ausentes.

#### **Linear**

A configuração padrão calcula o estimador linear.

#### **Executar o**

Calcula o estimador de execução.

#### **Quadrático**

Calcula o estimador quadrático.

### **Processo de iteração**

Fornecer configurações para especificar o estimador de iteração e o ajuste de erro padrão.

### **Modelo de efeitos fixos**

Quando selecionado, um modelo de efeitos fixos é empregado e as opções de estimativa de iteração e ajuste de erro padrão não estão disponíveis.

### **Modelo de efeitos aleatórios**

Quando selecionado, um modelo de efeitos aleatórios é empregado e as seguintes configurações estão disponíveis.

#### **Estimador**

Fornecer configurações para especificar o estimador de iteração

#### **Máxima verossimilhança restrita (REML)**

A configuração padrão aplica o método iterativo e calcula o estimador de máxima verossimilhança restrito.

#### **Máxima verossimilhança (ML)**

Aplica o método iterativo e calcula o estimador de máxima verossimilhança.

#### **Bayes empírico**

Aplica o método iterativo e calcula o estimador Bayes empírico.

#### **Hedges**

Aplica o método não iterativo e calcula o estimador de Hedges

#### **Hunter-Schmidt**

Aplica o método não iterativo e calcula o estimador de Hunter-Schmidt.

#### **DerSimonian-Laird**

Aplica o método não iterativo e calcula o estimador DerSimonian-Laird.

#### **Sidik-Jonkman**

Aplica o método não iterativo e calcula o estimador de Sidik-Jonkman.

#### **Ajuste de erro padrão**

Fornecer configurações para controlar se o ajuste de erro padrão de Knapp-Hartung deve ser aplicado às iterações do algoritmo de trim-and-fill.

#### **Nenhum ajuste**

A configuração padrão não aplica o ajustamento.

#### **Aplicar o ajuste de Knapp-Hartung**

Aplica o método de ajustamento Knapp-Hartung.

#### **Aplicar o ajuste truncado de Knapp-Hartung**

Aplica o método de ajuste de Knapp-Hartung e trunca o valor se for menor que 1 ao estimar a matriz de variância-covariância.

## **Definindo configurações de trim-and-fill de Binário de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Dados Brutos...**

2. No diálogo **Binário de metanálise**, clique em **Trim-and-Fill**.

3. Selecione e defina as configurações apropriadas de aparar e preencher.

4. Clique **Continuar**.

## **Binário de metanálise: imprimir**

O diálogo **Imprimir** fornece configurações para controlar saídas de tabela para metanálise com resultados binários em dados brutos que são fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### **Homogeneidade/Heterogeneidade**

Fornecer configurações para controlar os testes de homogeneidade e heterogeneidade.

#### **Teste de homogeneidade**

Quando selecionado, o teste de homogeneidade correspondente é fornecido na saída.

### **Medidas de heterogeneidade**

Quando selecionadas, as medidas de heterogeneidade são fornecidas na saída.

### **Tamanhos do efeito**

Fornecer as configurações de tamanho do efeito a seguir.

#### **Estudos individuais**

Controla a exibição de estudos individuais. Quando selecionado, a saída correspondente é fornecida.

#### **Tamanhos de efeito acumulativo**

Controla a exibição de análise acumulativa. Quando selecionado, a saída correspondente é fornecida. A configuração está disponível apenas quando uma variável de **Análise acumulativa** é selecionada no diálogo **Análise**.

#### **Intervalo de predição sob o modelo de efeitos aleatórios**

Controla a exibição do intervalo de predição. Quando selecionado, a saída correspondente é fornecida. A configuração está disponível somente quando um modelo de efeitos aleatórios é especificado.

### **Exibir estatísticas exponenciadas**

Controla a inclusão das estatísticas de forma exponencial. Quando a configuração é selecionada, as estatísticas de forma exponencial, incluindo o tamanho do efeito exponencial e os limites de intervalo de confiança são de saída. O processo aplica-se a ambas as tabelas de estimação e predição de tamanho do efeito necessário e opcional. A configuração está disponível quando o **Tamanho do Efeito** é especificado como **Razão de Chances de Logs**, **Razão de Chances de Logs de Peto** ou **Razão de Risco de Log**.

## **Definindo configurações de impressão de Binário de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Dados Brutos...**

2. No diálogo **Binário de metanálise**, clique em **Imprimir**.

3. Selecione e defina as configurações apropriadas de impressão.

4. Clique **Continuar**.

## **Binário de metanálise: salvar**

O diálogo **Salvar** fornece configurações para salvar as estatísticas estimadas no conjunto de dados ativo para metanálise com resultados binários em dados brutos fornecidos no conjunto de dados ativo para a estimação do tamanho do efeito.

### **Estudos individuais**

Fornecer opções para salvar o tamanho do efeito estimado.

#### **Tamanho do efeito individual**

Salva o tamanho do efeito estimado.

#### **Tamanho do efeito individual (forma exponenciada)**

Salva o tamanho do efeito estimado na forma exponencial.

#### **Erro padrão**

Salva o erro padrão estimado do tamanho do efeito.

#### **Limite inferior de intervalo de confiança**

Salva o limite inferior de intervalo de confiança estimado do tamanho do efeito.

#### **Limite superior de intervalo de confiança**

Salva o limite superior de intervalo de confiança estimado do tamanho do efeito.

#### **Limite inferior do intervalo de confiança (forma exponencial)**

Salva o limite inferior de intervalo de confiança estimado do tamanho do efeito na forma exponencial.

**Limite superior do intervalo de confiança (forma exponencial)**

Salva o limite superior de intervalo de confiança estimado do tamanho do efeito na forma exponencial.

**Valor p**

Salva o valor  $p$  estimado do tamanho do efeito.

**Ponderação de estudo**

Salva o peso do estudo estimado.

**Porcentagem de ponderação do estudo**

Salva o peso do estudo normalizado como uma porcentagem.

**Definindo configurações de salvamento de Binário de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Dados Brutos...**

2. No diálogo **Binário de metanálise**, clique em **Salvar**.

3. Selecione e defina as configurações apropriadas de salvamento das estatísticas estimadas.

4. Clique **Continuar**.

**Binário de metanálise: gráfico**

O diálogo **Gráfico** fornece configurações para os tipos de gráfico a seguir:

- “Gráfico de floresta” na página 77
- “Gráfico de floresta acumulativo” na página 78
- “Gráfico de bolhas” na página 79
- “Gráfico de funil” na página 80
- “Gráfico de Galbraith” na página 81
- “Gráfico de L'Abbé” na página 82

**Tamanho do Efeito Binário de Metanálise**

O procedimento de Tamanho do efeito binário de metanálise executa a metanálise, com resultados binários quando os dados pré-calculados de tamanho do efeito são fornecidos no conjunto de dados ativo.

**Exemplo**

Vários estudos de pesquisa foram realizados no passado para investigar um medicamento da moda, porém questionável, para ajudar no tratamento de diabetes tipo II. O medicamento oral foi alegado ser capaz de reduzir o nível de glicose no sangue após as refeições. Os dados foram coletados em diferentes locais de pesquisa de 1979 a 1986.

Um investigador principal gostaria de fazer inferência estatística sobre o efeito do medicamento oral. Devido ao fato de que os dados foram gerados a partir de diferentes estudos, ela propôs a ideia de sintetizar os resultados entre os estudos para alcançar uma compreensão geral do efeito e identificar as fontes subjacentes de variação nos resultados.

**Estatísticas**

Intervalo de confiança, Razão de Chances de Log, Razão de Chances de Log de Peto, Razão de Risco de Log, diferença de risco, efeitos aleatórios, efeitos fixos, variância inversa, Mantel-Haenszel, iterações, fracionamento de passos, convergência, estatísticas acumulativas, tamanho do efeito acumulativo, máxima verossimilhança restrita, REML, máxima verossimilhança, ML, Bayes Empírico, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, Teste de Egger, Teste de Harbord, Teste de Peters, interceptação na regressão, parâmetro de dispersão, homogeneidade, heterogeneidade, estatística exponencial, erro padrão, valor  $p$ , ponderação de estudo.



## Obtendo uma análise de Tamanho do efeito binário de metanálise

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Tamanho do Efeito Pré-calculado...**

2. Selecione uma variável de **Tamanho do efeito** que denote o tamanho do efeito. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).
3. Selecione **Desvio padrão** para especificar o desvio padrão de tamanho do efeito ou **Variação** para especificar a variação de tamanho do efeito e, em seguida, selecione uma variável para representar o desvio/a variação padrão para o tamanho do efeito.
4. Opcionalmente, selecione as variáveis **ID do Estudo** e/ou **Rótulo do Estudo**. A variável **ID do Estudo** selecionada não pode ser igual à variável **Rótulo do Estudo** selecionada.
5. Opcionalmente, selecione a configuração **Tamanho do Efeito**. As opções disponíveis são **Razão de chances de log**, **Razão de chances de log de Peto**, **Razão de risco de log** e **Diferença de risco**.
6. Opcionalmente, selecione uma configuração **Modelo**. Quando as configurações **Aparar e Preencher** são ativadas, elas também controlam o modelo usado, agrupando a análise aparar e preencher. Quando as configurações **Viés** são ativadas, elas também controlam o modelo usado pelo teste baseado em regressão.

### Efeitos aleatórios

A configuração padrão desenvolve o modelo de efeitos aleatórios.

### Efeitos fixos

Desenvolve o modelo de efeitos fixos.

7. Opcionalmente, é possível:

- Clique em **Critérios...** para especificar os critérios gerais.
- Clique em **Análise** para especificar o subgrupo e a análise acumulativa.
- Clique em **Inferência** para especificar os métodos de estimação.
- Clique em **Contraste** para controlar o teste de contraste.
- Clique em **Viés** para acessar o viés de publicação, realizando o teste baseado em regressão da EGGER.
- Clique em **Aparar e Preencher** para implementar a análise aparar e preencher do viés da publicação.
- Clique em **Imprimir** para controlar as saídas de tabela.
- Clique em **Salvar** para salvar as estatísticas estimadas para o conjunto de dados ativo.
- Clique em **Gráfico** para especificar quais gráficos incluir na saída.

8. Clique em **OK**.

## Tamanho do efeito binário de metanálise: critérios

O diálogo **Critérios** fornece configurações para especificar critérios para metanálise com resultados binários quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### Intervalo de confiança

A configuração opcional especifica o nível de confiança. O valor deve ser um valor numérico entre 0 e 100. A configuração padrão é 95.

### Escopo de dados omissos

As configurações opcionais controlam como o procedimento manipula dados ausentes.

### Excluir análise de casos por análise

A configuração padrão inclui todos os casos com dados suficientes sobre as variáveis usadas em cada análise específica.

### **Excluir casos de listwise**

Inclui todos os casos com dados suficientes sobre todas as variáveis usadas em todas as análises especificadas pelo procedimento.

### **Valores ausentes do usuário**

As configurações opcionais controlam como os valores ausentes do usuário são tratados.

#### **Exclui**

A configuração padrão trata os valores ausentes do usuário como válidos.

#### **Inclui**

Ignora as designações de valor omissos do usuário e trata os valores omissos do usuário como válidos.

### **Iterações**

#### **Iterações máximas**

A configuração opcional especifica o número máximo de iterações nos métodos iterativos. A configuração está disponível quando métodos iterativos são usados. O valor deve ser um único número inteiro positivo. O valor-padrão é 100. Um valor 0 significa que não são realizadas iterações.

#### **Redução máxima da etapa pela metade**

A configuração opcional especifica a redução máxima pela metade da etapa nos métodos iterativos. A configuração está disponível quando métodos iterativos são usados. O valor deve ser um único número inteiro positivo. O valor padrão é 5 segundos. Um valor de 0 significa que a redução da etapa pela metade não é aplicada.

#### **Convergência**

A configuração opcional especifica a tolerância de convergência. O valor deve ser um valor positivo único. A configuração está disponível quando métodos iterativos são usados. O valor padrão é 1E-6.

## **Definindo critérios de Tamanho do efeito binário de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Tamanho do Efeito Pré-calculado...**

2. No diálogo **Tamanho do efeito binário de metanálise**, clique em **Crítérios**.

3. Selecione e defina as configurações apropriadas de critérios.

4. Clique **Continuar**.

## **Tamanho do efeito binário de metanálise: análise**

O diálogo **Análise** fornece configurações para especificar análise de subgrupo e acumulativa para metanálise com resultados binários quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### **Análise de subgrupo**

Selecione uma variável que evoque a análise de subgrupo. A variável não pode ser a mesma que a variável definida para **Análise acumulativa**.

### **Análise acumulativa**

Selecione uma variável que evoca a análise acumulativa e sobre a qual a meta-análise acumulativa é conduzida. A variável não pode ser a mesma que a variável definida para **Análise de subgrupo**. Quando **Crescente** é selecionado, a análise acumulativa é baseada na variável especificada em ordem crescente. Quando **Decrescente** é selecionado, a análise acumulativa é baseada na variável especificada em ordem decrescente.

### **Estatísticas acumulativas**

Fornece opções para salvar o tamanho do efeito geral acumulativo estimado. As configurações estão disponíveis apenas quando uma variável de **Análise acumulativa** é selecionada.

**Tamanho do efeito acumulativo**

Salva o tamanho do efeito geral acumulativo estimado.

**Tamanho do efeito acumulativo (forma exponenciada)**

Salva o tamanho do efeito geral acumulativo estimado na forma exponencial.

**Erro padrão**

Salva o erro padrão estimado do tamanho do efeito geral acumulativo.

**Limite inferior de intervalo de confiança**

Salva o limite inferior do intervalo de confiança estimado do tamanho do efeito geral acumulativo.

**Limite superior de intervalo de confiança**

Salva o limite superior do intervalo de confiança estimado do tamanho do efeito geral acumulativo.

**Limite inferior do intervalo de confiança (forma exponencial)**

Salva o limite inferior do intervalo de confiança estimado do tamanho do efeito geral acumulativo na forma exponencial.

**Limite superior do intervalo de confiança (forma exponencial)**

Salva o limite superior do intervalo de confiança estimado do tamanho do efeito geral acumulativo na forma exponencial.

**Valor p**

Salva o valor  $p$  estimado do tamanho do efeito geral acumulativo.

**Destino**

Fornece opções para especificar um conjunto de dados ou arquivo de dados salvo. Ao selecionar **Conjunto de dados**, é possível especificar um novo nome de conjunto de dados ou manter o nome padrão. Ao selecionar **Arquivo de dados**, clique em **Navegar...** para selecionar o nome e o local de um arquivo salvo.

**Definindo configurações de análise de Tamanho do efeito binário de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Tamanho do Efeito Pré-calculado...**

2. No diálogo **Tamanho do efeito binário de metanálise**, clique em **Análise**.

3. Selecione e defina as configurações apropriadas de análise.

4. Clique **Continuar**.

**Tamanho do efeito binário de metanálise: inferência**

O diálogo **Inferência** fornece configurações para especificar os métodos de estimação para metanálise com resultados binários quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

**Estimador**

Fornece configurações para especificar o estimador.

**Notas:**

- Quando as configurações de Aparar e Preencher são especificadas, também controla o estimador usado pelo agrupamento na análise de aparar e preencher.
- Quando as configurações de Viés são especificadas, também controla o estimador usado pelo teste baseado em regressão.

**Máxima verossimilhança restrita (REML)**

A configuração padrão aplica o método iterativo e calcula o estimador de máxima verossimilhança restrito.

**Máxima verossimilhança (ML)**

Aplica o método iterativo e calcula o estimador de máxima verossimilhança.

**Bayes empírico**

Aplica o método iterativo e calcula o estimador Bayes empírico.

**Hedges**

Aplica o método não iterativo e calcula o estimador de Hedges.

**Hunter-Schmidt**

Aplica o método não iterativo e calcula o estimador de Hunter-Schmidt.

**DerSimonian-Laird**

Aplica o método não iterativo e calcula o estimador DerSimonian-Laird.

**Sidik-Jonkman**

Aplica o método não iterativo e calcula o estimador de Sidik-Jonkman.

**Ajuste de erro padrão**

Fornece configurações que controlam se o ajuste de erro padrão de Knapp-Hartung deve ser aplicado.

**Notas:**

- Quando as configurações de corte e preenchimento são especificadas, também controla o ajuste de erro padrão que é usado pelo agrupamento na análise de corte e preenchimento.
- Quando as configurações de Biais são especificadas, também controla o ajuste de erro padrão usado pelo teste baseado em regressão.

**Nenhum ajuste**

A configuração padrão não aplica o ajustamento.

**Aplicar o ajuste de Knapp-Hartung**

Aplica o método de ajustamento Knapp-Hartung.

**Aplicar o ajuste truncado de Knapp-Hartung**

Aplica o método de ajuste de Knapp-Hartung e trunca o valor se for menor que 1 ao estimar a matriz de variância-covariância.

**Definindo configurações de inferência de Tamanho do efeito binário de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Tamanho do Efeito Pré-calculado...**

2. No diálogo **Tamanho do efeito binário de metanálise**, clique em **Inferência**.

3. Selecione e defina as configurações apropriadas de inferência.

4. Clique **Continuar**.

**Tamanho do efeito binário de metanálise: contraste**

O diálogo **Contraste** fornece configurações para controlar o teste de contraste para metanálise com resultados binários quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

**Variáveis**

A lista exibe todas as variáveis de conjunto de dados disponíveis. Selecione variáveis na lista e mova-as para a lista de **Contrastes**.

**Contrastes**

A lista identifica os coeficientes armazenados como variáveis no conjunto de dados ativo. Várias variáveis são permitidas. Variáveis de sequência de caracteres não são suportadas.

**Valores de coeficientes de entrada do usuário**

Fornece configurações para especificar coeficientes de um contraste fornecido pelo usuário. Apenas valores numéricos são permitidos. Para formular um teste de contraste válido, o número de valores especificados deve corresponder ao dos estudos válidos.

**Exibir estatísticas exponenciadas**

Controla a inclusão das estatísticas de forma exponencial. Quando a configuração é selecionada, a saída inclui as estatísticas de forma exponencial, incluindo o tamanho do efeito exponencial e os

limites do intervalo de confiança. A configuração está disponível quando o **Tamanho do Efeito** é especificado como **Razão de Chances de Logs**, **Razão de Chances de Logs de Peto** ou **Razão de Risco de Log**.

## Definindo configurações de contraste de Tamanho do efeito binário de metanálise

1. No menu, escolha:

**Analisar > Metanálise > Resultados Binários > Tamanho do Efeito Pré-calculado...**

2. No diálogo **Tamanho do efeito binário de metanálise**, clique em **Contraste**.

3. Selecione e defina as configurações apropriadas de contraste.

4. Clique **Continuar**.

## Tamanho do efeito binário de metanálise: viés

O diálogo **Viés** fornece configurações para ativar o viés de publicação realizando testes baseados em regressão para metanálise, com resultados binários quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### Teste baseado em regressão de Egger

A seleção dessa configuração ativa o viés de publicação realizando o teste baseado em regressão de Eggers.

### Variáveis

A lista fornece todas as variáveis de conjunto de dados disponíveis.

### Covariáveis

Variáveis selecionadas na lista de **Variáveis** são tratadas como covariáveis. Várias covariáveis são permitidas.

### Fatores

Variáveis selecionadas na lista de **Variáveis** são tratadas como fatores. Vários fatores são permitidos.

### Incluir interceptação na regressão

Controla o termo de intercepto no teste baseado em regressão.

### Incluir parâmetro de dispersão em modelo de efeitos fixos

Controla a configuração de modelo multiplicativo e apresenta o parâmetro de dispersão multiplicativa para a análise. A configuração está disponível apenas quando um modelo de efeitos fixos é selecionado.

### Estimar estatísticas com base na distribuição t

Controla a distribuição usada nos testes baseados em regressão. A configuração é ativada por padrão e estima as estatísticas com base em *t*-distribution. Se a configuração não for selecionada, as estatísticas serão estimadas com base na distribuição normal.

## Definindo configurações de viés de Tamanho do efeito binário de metanálise

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Tamanho do Efeito Pré-calculado...**

2. No diálogo **Tamanho do efeito binário de metanálise**, clique em **Viés**.

3. Selecione e defina as configurações apropriadas de viés.

4. Clique **Continuar**.

## Tamanho do efeito binário de metanálise: trim-and-fill

O diálogo **Trim-and-Fill** fornece configurações para implementar a análise de trim-and-fill de viés de publicação para metanálise com resultados binários quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

**Estimar o número de estudos omissos**

Controla a análise de trim-and-fill de viés de publicação. A seleção dessa configuração habilita as outras configurações de diálogo.

**Estudos colaterais para imputação**

Fornecer opções para especificar o lado do gráfico de funil no qual os estudos omissos são imputados.

**Determinado pelas inclinações do teste de Egger**

A configuração padrão determina o lado com base na inclinação estimada do teste de Egger.

**Esquerda**

Imputa o lado esquerdo do gráfico de funil.

**Direita**

Imputa o lado direito do gráfico de funil.

**Método**

Especifica o método para estimar o número de estudos ausentes.

**Linear**

A configuração padrão calcula o estimador linear.

**Executar o**

Calcula o estimador de execução.

**Quadrático**

Calcula o estimador quadrático.

**Processo de iteração**

Fornecer configurações para especificar o estimador de iteração e o ajuste de erro padrão.

**Modelo de efeitos fixos**

Quando selecionado, um modelo de efeitos fixos é empregado e as opções de estimativa de iteração e ajuste de erro padrão não estão disponíveis.

**Modelo de efeitos aleatórios**

Quando selecionado, um modelo de efeitos aleatórios é empregado e as seguintes configurações estão disponíveis.

**Estimador**

Fornecer configurações para especificar o estimador de iteração

**Máxima verossimilhança restrita (REML)**

A configuração padrão aplica o método iterativo e calcula o estimador de máxima verossimilhança restrito.

**Máxima verossimilhança (ML)**

Aplica o método iterativo e calcula o estimador de máxima verossimilhança.

**Bayes empírico**

Aplica o método iterativo e calcula o estimador Bayes empírico.

**Hedges**

Aplica o método não iterativo e calcula o estimador de Hedges

**Hunter-Schmidt**

Aplica o método não iterativo e calcula o estimador de Hunter-Schmidt.

**DerSimonian-Laird**

Aplica o método não iterativo e calcula o estimador DerSimonian-Laird.

**Sidik-Jonkman**

Aplica o método não iterativo e calcula o estimador de Sidik-Jonkman.

**Ajuste de erro padrão**

Fornecer configurações para controlar se o ajuste de erro padrão de Knapp-Hartung deve ser aplicado às iterações do algoritmo de trim-and-fill.

**Nenhum ajuste**

A configuração padrão não aplica o ajustamento.

### **Aplicar o ajuste de Knapp-Hartung**

Aplica o método de ajustamento Knapp-Hartung.

### **Aplicar o ajuste truncado de Knapp-Hartung**

Aplica o método de ajuste de Knapp-Hartung e trunca o valor se for menor que 1 ao estimar a matriz de variância-covariância.

## **Definindo configurações de trim-and-fill de Tamanho do efeito binário de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Tamanho do Efeito Pré-calculado...**

2. No diálogo **Tamanho do efeito binário de metanálise**, clique em **Trim-and-Fill**.

3. Selecione e defina as configurações apropriadas de aparar e preencher.

4. Clique **Continuar**.

## **Tamanho do efeito binário de metanálise: imprimir**

O diálogo **Imprimir** fornece configurações para controlar saídas de tabela para metanálise com resultados binários quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### **Homogeneidade/Heterogeneidade**

Fornece configurações para controlar os testes de homogeneidade e heterogeneidade.

#### **Teste de homogeneidade**

Quando selecionado, o teste de homogeneidade correspondente é fornecido na saída.

#### **Medidas de heterogeneidade**

Quando selecionadas, as medidas de heterogeneidade são fornecidas na saída.

### **Tamanhos do efeito**

Fornece as configurações de tamanho do efeito a seguir.

#### **Estudos individuais**

Controla a exibição de estudos individuais. Quando selecionado, a saída correspondente é fornecida.

#### **Tamanhos de efeito acumulativo**

Controla a exibição de análise acumulativa. Quando selecionado, a saída correspondente é fornecida. A configuração está disponível apenas quando uma variável de **Análise acumulativa** é selecionada no diálogo **Análise**.

#### **Intervalo de predição sob o modelo de efeitos aleatórios**

Controla a exibição do intervalo de predição. Quando selecionado, a saída correspondente é fornecida. A configuração está disponível somente quando um modelo de efeitos aleatórios é especificado.

### **Exibir estatísticas exponenciadas**

Controla a inclusão das estatísticas de forma exponencial. Quando a configuração é selecionada, as estatísticas de forma exponencial, incluindo o tamanho do efeito exponencial e os limites de intervalo de confiança são de saída. O processo aplica-se a ambas as tabelas de estimação e predição de tamanho do efeito necessário e opcional. A configuração está disponível quando o **Tamanho do Efeito** é especificado como **Razão de Chances de Logs**, **Razão de Chances de Logs de Peto** ou **Razão de Risco de Log**.

## **Definindo configurações de impressão de Tamanho do efeito binário de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Binários > Tamanho do Efeito Pré-calculado...**

2. No diálogo **Tamanho do efeito binário de metanálise**, clique em **Imprimir**.

3. Selecione e defina as configurações apropriadas de impressão.
4. Clique **Continuar**.

## Tamanho do efeito binário de metanálise: salvar

O diálogo **Salvar** fornece configurações para salvar as estatísticas estimadas no conjunto de dados ativo para metanálise com resultados binários quando os dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### Estudos individuais

Fornecer opções para salvar o tamanho do efeito estimado.

#### Tamanho do efeito individual (forma exponenciada)

Salva o tamanho do efeito estimado na forma exponencial.

#### Erro padrão

Salva o erro padrão estimado do tamanho do efeito.

#### Limite inferior de intervalo de confiança

Salva o limite inferior de intervalo de confiança estimado do tamanho do efeito.

#### Limite superior de intervalo de confiança

Salva o limite superior de intervalo de confiança estimado do tamanho do efeito.

#### Limite inferior do intervalo de confiança (forma exponencial)

Salva o limite inferior de intervalo de confiança estimado do tamanho do efeito na forma exponencial.

#### Limite superior do intervalo de confiança (forma exponencial)

Salva o limite superior de intervalo de confiança estimado do tamanho do efeito na forma exponencial.

#### Valor p

Salva o valor  $p$  estimado do tamanho do efeito.

#### Ponderação de estudo

Salva o peso do estudo estimado.

#### Porcentagem de ponderação do estudo

Salva o peso do estudo normalizado como uma porcentagem.

## Definindo configurações de salvamento do Tamanho do efeito binário de metanálise

1. No menu, escolha:  
**Analisar > Metanálise > Resultados Binários > Tamanho do Efeito Pré-calculado...**
2. No diálogo **Tamanho do efeito binário de metanálise**, clique em **Salvar**.
3. Selecione e defina as configurações apropriadas de salvamento das estatísticas estimadas.
4. Clique **Continuar**.

## Tamanho do efeito binário de metanálise: criar gráfico

O diálogo **Gráfico** fornece configurações para os tipos de gráfico a seguir:

- [“Gráfico de floresta” na página 77](#)
- [“Gráfico de floresta acumulativo” na página 78](#)
- [“Gráfico de bolhas” na página 79](#)
- [“Gráfico de funil” na página 80](#)
- [“Gráfico de Galbraith” na página 81](#)
- [“Gráfico de L'Abbé” na página 82](#)



## Regressão de Metanálise

O procedimento de Regressão de metanálise executa a análise de metarregressão.

### Exemplo

Vários estudos de pesquisa foram realizados no passado para investigar um medicamento da moda, porém questionável, para ajudar no tratamento de diabetes tipo II. O medicamento oral foi alegado ser capaz de reduzir o nível de glicose no sangue após as refeições. Os dados foram coletados em diferentes locais de pesquisa de 1979 a 1986.

Um investigador principal gostaria de fazer inferência estatística sobre o efeito do medicamento oral. Devido ao fato de que os dados foram gerados a partir de diferentes estudos, ela propôs a ideia de sintetizar os resultados entre os estudos para alcançar uma compreensão geral do efeito e identificar as fontes subjacentes de variação nos resultados.

### Estatísticas

Nível de confiança, método iterativo, redução de etapa pela metade, tolerância de convergência, médias da amostra, variância da amostra, desvio padrão, tamanho do efeito estimado, método de estimação, teste baseado em regressão, modelo de efeitos aleatórios, modelo de efeitos fixos, parâmetro de dispersão, estimador de máxima verossimilhança restrita, estimador de Bayes empírico, estimador de Hedges, estimador de Hunter-Schmidt, estimador de DerSimonian-Laird, estimador de Sidik-Jonkman, ajuste de erro padrão de Knapp-Hartung, ajuste de erro padrão de Knapp-Hartung truncado, coeficientes, teste de coeficiente de modelo, estatísticas exponenciadas.

## Obtendo uma análise de Regressão de metanálise

1. A partir dos menus, escolha:

**Analisar > Metanálise > Metarregressão**

2. Selecione uma única variável de **Tamanho do efeito** que denote o tamanho do efeito. A variável selecionada deve ser numérica (variáveis de sequência de caracteres não são suportadas).
3. Selecione uma das configurações a seguir e, em seguida, selecione uma variável numérica única correspondente:

#### Erro padrão

Selecione uma variável que especifique o erro padrão que é convertida para ponderação. Essa é a configuração padrão.

#### Variância

Selecione uma variável que especifique a variação que é convertida para ponderação.

#### Peso

Selecione uma variável que especifique a ponderação.

4. Opcionalmente, inclua variáveis fator na lista de **Fatores**. Para cada variável fator selecionada um valor opcional pode ser especificado para designar **Últimas categorias** customizadas. É possível clicar em **Reconfigurar** para restaurar os valores de **Últimas categorias** para as suas configurações padrão.

**Nota:** Quando não houver casos que correspondam aos valores de **Últimas categorias** especificados, os últimos valores ocorridos serão tratados como as últimas categorias.

5. Opcionalmente, selecione variáveis covariáveis numéricas.
6. Opcionalmente, selecione uma configuração **Modelo**.

#### Efeitos aleatórios

A configuração padrão desenvolve o modelo de efeitos aleatórios.

#### Efeitos fixos

Desenvolve o modelo de efeitos fixos. É possível selecionar, opcionalmente, a configuração **Incluir parâmetro de dispersão**.

7. Opcionalmente, é possível:

- Clique em **Crítérios...** para especificar os critérios gerais.
- Clique em **Inferência** para especificar os métodos de estimação.

- Clique em **Imprimir** para controlar as saídas de tabela.
  - Clique em **Salvar** para prever e salvar as estatísticas estimadas para o conjunto de dados ativo.
  - Clique em **Gráfico** para especificar quais gráficos incluir na saída.
8. Clique em **OK**.

## Regressão de metanálise: critérios

O diálogo **Críticos** fornece configurações para especificar critérios para regressão de metanálise.

### Intervalo de confiança

A configuração opcional especifica o nível de confiança. O valor deve ser um valor numérico entre 0 e 100. A configuração padrão é 95.

### Valores ausentes do usuário

As configurações opcionais controlam como os valores ausentes do usuário são tratados.

#### Excluído

A configuração padrão trata os valores ausentes do usuário como válidos.

#### Incluído

Ignora as designações de valor omissos do usuário e trata os valores omissos do usuário como válidos.

### Iterações

#### Iterações máximas

A configuração opcional especifica o número máximo de iterações nos métodos iterativos. A configuração está disponível quando métodos iterativos são usados. O valor deve ser um único número inteiro positivo. O valor-padrão é 100. Um valor 0 significa que não são realizadas iterações.

#### Redução máxima da etapa pela metade

A configuração opcional especifica a redução máxima pela metade da etapa nos métodos iterativos. A configuração está disponível quando métodos iterativos são usados. O valor deve ser um único número inteiro positivo. O valor padrão é 5 segundos. Um valor de 0 significa que a redução da etapa pela metade não é aplicada.

#### Convergência

A configuração opcional especifica a tolerância de convergência. O valor deve ser um valor positivo único. A configuração está disponível quando métodos iterativos são usados. O valor padrão é 1E-6.

## Definindo critérios de Regressão de metanálise

1. A partir dos menus, escolha:

**Analisar > Metanálise > Metarregressão**

2. No diálogo **Regressão de metanálise**, clique em **Críticos**.
3. Selecione e defina as configurações apropriadas de critérios.
4. Clique **Continuar**.

## Regressão de metanálise: inferência

O diálogo **Inferência** fornece configurações para especificar os métodos de estimação para regressão de metanálise.

### Incluir interceptação na regressão

A configuração é ativada por padrão.

### Estimar estatísticas com base na distribuição t

Controla a distribuição usada nos testes baseados em regressão. A configuração é ativada por padrão e estima as estatísticas com base em *t*-distribution. Se a configuração não for selecionada, as estatísticas serão estimadas com base na distribuição normal.

## **Estimador**

Fornecer configurações para especificar o estimador.

### **Máxima verossimilhança restrita (REML)**

A configuração padrão aplica o método iterativo e calcula o estimador de máxima verossimilhança restrito.

### **Máxima verossimilhança (ML)**

Aplica o método iterativo e calcula o estimador de máxima verossimilhança.

### **Bayes empírico**

Aplica o método iterativo e calcula o estimador Bayes empírico.

### **Hedges**

Aplica o método não iterativo e calcula o estimador de Hedges.

### **Hunter-Schmidt**

Aplica o método não iterativo e calcula o estimador de Hunter-Schmidt.

### **DerSimonian-Laird**

Aplica o método não iterativo e calcula o estimador DerSimonian-Laird.

### **Sidik-Jonkman**

Aplica o método não iterativo e calcula o estimador de Sidik-Jonkman.

## **Ajuste de erro padrão**

Fornecer configurações que controlam se o ajuste de erro padrão de Knapp-Hartung deve ser aplicado.

### **Nenhum ajuste**

A configuração padrão não aplica o ajustamento.

### **Aplicar o ajuste de Knapp-Hartung**

Aplica o método de ajustamento Knapp-Hartung.

### **Aplicar o ajuste truncado de Knapp-Hartung**

Aplica o método de ajuste de Knapp-Hartung e trunca o valor se for menor que 1 ao estimar a matriz de variância-covariância.

## **Definindo configurações de inferência de Regressão de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Metarregressão**

2. No diálogo **Regressão de metanálise**, clique em **Inferência**.

3. Selecione e defina as configurações apropriadas de inferência.

4. Clique **Continuar**.

## **Regressão de metanálise: imprimir**

O diálogo **Imprimir** fornece configurações para controlar saídas de tabela para regressão de metanálise.

### **Teste de coeficiente de modelo**

Controla o teste de coeficientes do modelo. A configuração não é especificada por padrão, o que suprime o teste. Quando a configuração é especificada, o teste é fornecido na saída.

### **Exibir estatísticas exponenciadas**

Controla as estimativas de parâmetro. A configuração não é especificada por padrão, o que suprime as estimativas de parâmetro. Quando a configuração é especificada, as estimativas de parâmetro são fornecidas na saída.

## **Definindo configurações de impressão de Regressão de metanálise**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Metarregressão**

2. No diálogo **Regressão de metanálise**, clique em **Imprimir**.

3. Selecione e defina as configurações apropriadas de impressão.
4. Clique **Continuar**.

## Regressão de metanálise: salvar

O diálogo **Salvar** fornece configurações para salvar as estatísticas estimadas no conjunto de dados ativo para regressão de metanálise.

### Estudos individuais

Fornecer opções para salvar o tamanho do efeito estimado.

#### Valores preditos

Prediz e salva os valores ajustados.

#### Erro padrão de valores preditos

Prediz e salva o erro padrão de valores ajustados.

#### Limite inferior de intervalo de confiança

Salva o limite inferior do intervalo de confiança estimado dos valores ajustados.

#### Limite superior de intervalo de confiança

Salva o limite superior do intervalo de confiança estimado dos valores ajustados.

#### Resíduos

Prediz e salva os resíduos.

#### Erro padrão de resíduos

Prediz e salva o erro padrão dos resíduos.

#### Alavancagens

Prediz e salva as alavancagens.

### Estatísticas acumulativas

Fornecer opções para salvar previsões lineares.

**Nota:** As configurações de estatísticas acumulativas são ignoradas quando **Efeitos fixos** são selecionados como a configuração **Modelo**.

#### Predições lineares fixas

Prediz e salva a previsão linear.

#### Erro padrão de previsões lineares fixas

Prediz e salva o erro padrão da previsão linear.

#### Melhores previsões lineares imparciais

Prediz e salva a melhor previsão linear imparcial dos efeitos aleatórios.

#### Erro padrão de BLUPs

Prediz e salva o erro padrão da melhor previsão linear imparcial dos efeitos aleatórios.

## Definindo configurações de salvamento de Regressão de metanálise

1. A partir dos menus, escolha:

**Analisar > Metanálise > Metarregressão**

2. No diálogo **Regressão de metanálise**, clique em **Salvar**.
3. Selecione e defina as configurações de salvamento apropriadas.
4. Clique **Continuar**.

## Regressão de metanálise: criar gráfico

O diálogo **Criar gráfico** fornece configurações para definir gráficos de bolhas. Para obter mais informações, consulte [“Gráfico de bolhas”](#) na página 79.

## Opções de plotagem de metanálise

As opções de gráfico a seguir estão disponíveis para os procedimentos de Metanálise e são acessíveis nos diálogos **Criar gráfico**.

### Gráfico de floresta

A guia **Gráfico de floresta** fornece configurações para controlar gráficos de plot de floresta que são exibidos na saída para metanálise com resultados contínuos e binários quando dados brutos e dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

#### Gráfico de floresta

A configuração opcional ativa e desativa a saída do gráfico de floresta. A configuração está desativada por padrão.

#### Exibir colunas

As configurações opcionais especificam as estatísticas estimadas que são tabuladas.

#### Tamanho do efeito

Inclui o tamanho de efeito estimado.

#### Erro padrão

Inclui o erro padrão do tamanho do efeito estimado.

#### Limites de intervalo de confiança

Inclui os limites inferior e superior do intervalo de confiança do tamanho do efeito estimado.

#### Valor $p$

Inclui o valor  $p$  do tamanho do efeito estimado para cada estudo individual.

#### Peso

Inclui a ponderação dos estudos primários.

#### Exibir forma exponenciada

**Nota:** Essa configuração se aplica apenas à meta-análise com resultados binários.

Controla a coluna do gráfico e a exibição do tamanho do efeito e do intervalo de confiança. Quando a configuração é selecionada, plota a razão de chances, a razão de chances de Peto ou a razão de risco e tabula as estatísticas de forma exponencial. Quando a configuração não estiver selecionada, plota o log de odds ratio, o log de odds ratio de Peto ou o log de risco e tabula as estatísticas transformadas em log.

#### Variáveis

Lista as variáveis de conjunto de dados disponíveis.

#### Colunas adicionais

Opcionalmente, selecione variáveis para exibir como colunas adicionais. A ordem da variável determina a ordem de exibição da coluna.

#### Classificar

Opcionalmente, especifique uma variável sobre a qual o gráfico de floresta é classificado. Quando uma variável é especificada, o gráfico de floresta é classificado em ordem **Crescente** (a configuração padrão). Quando **decrecente** é selecionado, o gráfico de floresta é classificado em ordem decrescente.

#### Posição da coluna do gráfico

Controla o posicionamento da coluna de gráfico.

#### Direita

A configuração padrão coloca a coluna de plotagem à direita das outras colunas da tabela.

#### Esquerda

Coloca a coluna de gráfico à esquerda das outras colunas da tabela.

### Linhas de referência

Controla as linhas de referência que são incluídas no gráfico de floresta. **Tamanho do efeito geral** inclui a linha para denotar o tamanho do efeito geral estimado. **Tamanho do efeito nulo** inclui a linha para denotar o tamanho do efeito nulo.

### Anotações

Controla a exibição das anotações.

#### Homogeneidade

Imprime as estatísticas de teste de homogeneidade.

#### Heterogeneidade

Imprime as estatísticas de teste de heterogeneidade.

#### Test

A configuração padrão coloca a coluna de plotagem à direita das outras colunas da tabela.

### Intervalo de safra

Especifica a faixa de corte. Quando a configuração é selecionada, dois valores numéricos devem ser especificados e satisfazer  $Upper\ bound > Lower\ bound$ .

Para metanálise, com resultados binários, quando a configuração **Tamanho do Efeito** é definida para **Diferença de Risco**, ela também deve atender o  $Limite\ Superior \leq 1$  e o  $Limite\ Inferior \geq -1$ . Quando a configuração **Exibir formato exponenciado** é selecionada, ela também deve satisfazer  $Limite\ inferior \geq 0$ .

## Definindo configurações de gráfico de floresta

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos ou Resultados Binários > Dados Brutos... ou Tamanho do Efeito Pré-Calculado...**

2. No diálogo **Metanálise**, clique em **Criar gráfico** e, em seguida, clique na guia **Gráfico de floresta**.

3. Selecione e defina as configurações apropriadas de gráfico de floresta.

4. Clique **Continuar**.

## Gráfico de floresta acumulativo

A guia **Gráfico de floresta acumulativo** fornece configurações para controlar gráficos de plot de floresta acumulativos, que são exibidos na saída para metanálise, com resultados contínuos e binários quando dados brutos e dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### Gráfico de floresta acumulativo

A configuração opcional ativa e desativa a saída do gráfico de floresta. Este serviço está desativado por padrão.

### Exibir colunas

As configurações opcionais especificam as estatísticas estimadas que são tabuladas.

#### Tamanho do efeito acumulativo

Inclui o tamanho do efeito geral acumulativo.

#### Erro padrão

Inclui o erro padrão do tamanho do efeito geral acumulativo.

#### Limites de intervalo de confiança

Inclui os limites inferior e superior do intervalo de confiança do tamanho do efeito geral cumulativo.

#### Valor P

Inclui o valor  $p$  do tamanho do efeito geral acumulativo para cada estudo individual.

### Exibir forma exponenciada

**Nota:** Essa configuração se aplica apenas à meta-análise com resultados binários.

Controla a coluna do gráfico e a exibição do tamanho do efeito e do intervalo de confiança. Quando a configuração é selecionada, ela traça a razão de chances, a razão de chances de Peto ou a razão de risco e tabula as estatísticas de forma exponencial. Quando a configuração não está habilitada, ela traça o log de razão de chances, o log de razão de chances de Peto ou o log de razão de riscos e tabula as estatísticas transformadas em log.

### Variáveis

Lista as variáveis de conjunto de dados disponíveis.

### Colunas adicionais

Opcionalmente, selecione variáveis para exibir como colunas adicionais. A ordem da variável determina a ordem de exibição da coluna.

### Posição da coluna do gráfico

Controla o posicionamento da coluna de gráfico.

#### Direita

A configuração padrão coloca a coluna de plotagem à direita das outras colunas da tabela.

#### Esquerda

Coloca a coluna de gráfico à esquerda das outras colunas da tabela.

### Intervalo de safra

Especifica a faixa de corte. Quando a configuração é selecionada, dois valores numéricos devem ser especificados e satisfazer  $\text{Upper bound} > \text{Lower bound}$ .

Para metanálise, com resultados binários, quando a configuração **Tamanho do Efeito** é definida para **Diferença de Risco**, ela também deve atender o  $\text{Limite Superior} \leq 1$  e o  $\text{Limite Inferior} \geq -1$ . Quando a configuração **Exibir formato exponenciado** é selecionada, ela também deve satisfazer  $\text{Limite inferior} \geq 0$ .

## Definindo configurações de gráfico de floresta acumulativo

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos ou Resultados Binários > Dados Brutos... ou Tamanho do Efeito Pré-Calculado...**

2. No diálogo **Metanálise**, clique em **Criar gráfico** e em seguida, clique na guia **Gráfico de floresta acumulativo**.

3. Selecione e defina as configurações de gráfico de floresta acumulativo apropriadas.

4. Clique **Continuar**.

## Gráfico de bolhas

A guia **Gráfico de bolhas** fornece configurações para controlar gráficos de plot de bolhas que são exibidos na saída para metanálise com resultados contínuos e binários, quando dados brutos e dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

**Nota:** O Gráfico de bolhas é o único tipo de gráfico disponível para análise de regressão de metanálise.

### Gráfico de bolhas

A configuração opcional ativa e desativa a saída do gráfico de bolhas. Este serviço está desativado por padrão.

**Nota:** Essa configuração não está disponível para análise de regressão de meta-análise.

### Variáveis

Lista as variáveis de conjunto de dados disponíveis.

### Preditores

Especificar variáveis que são tratadas como preditores contínuos no eixo X. Várias variáveis são permitidas com cada uma criando um gráfico de bolhas separado.

**Nota:** Para análise de regressão de metanálise, esta configuração é rotulada como **Preditores de gráfico de bolhas** e permite selecionar covariáveis para as quais criar gráficos de bolhas.

### **Preditores centrados na média**

A configuração opcional controla o preditor contínuo no eixo X. A configuração está desativada por padrão, o que não centraliza o preditor. Quando a configuração é ativada, o preditor é centrado na média.

### **Rótulo**

Opcionalmente, especifique uma variável sobre a qual o gráfico de bolhas é rotulado. A lista **Colocação** fornece opções de colocação automática de rótulo (**Automático**) ou de colocação **À direita, À esquerda, Na parte superior e Na parte inferior**.

### **Desenhar bolhas proporcionais aos pesos**

Controla como as bolhas são desenhadas em relação às ponderações. A configuração é ativada por padrão, que desenha as bolhas proporcionais aos pesos. Quando a configuração está desativada, todas as bolhas são desenhadas com o mesmo tamanho.

### **Exibir linha ajustada**

Controla a exibição da linha de regressão ajustada. A configuração é ativada por padrão, o que inclui a linha de regressão ajustada no gráfico de bolhas. Quando a configuração está desativada, a linha de regressão ajustada não é incluída no gráfico de bolhas.

### **Exibir limites de intervalo de confiança**

Controla a exibição de limites de intervalo de confiança. A configuração é ativada por padrão, o que inclui os limites do intervalo de confiança no gráfico de bolhas. Quando a configuração está desativada, os limites do intervalo de confiança não são incluídos no gráfico de bolhas.

### **Intervalo de eixo X**

Especifica a amplitude do gráfico do eixo X. Quando a configuração é ativada, dois valores numéricos devem ser especificados e atender o **Limite Superior > Limite Inferior**. Observe que os valores especificados se aplicam a todos os gráficos de bolhas.

### **Intervalo de eixo Y**

Especifica a amplitude do gráfico do eixo Y. Quando a configuração é ativada, dois valores numéricos devem ser especificados e atender o **Limite Superior > Limite Inferior**. Observe que os valores especificados se aplicam a todos os gráficos de bolhas.

Para metanálise, com resultados binários, quando a configuração **Tamanho do Efeito** é definida para **Diferença de Risco**, ela também deve atender o **Limite Superior  $\leq 1$  e o Limite Inferior  $\geq -1$** .

## **Definindo configurações de gráfico de bolhas**

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos ou Resultados Binários > Dados Brutos... ou Tamanho do Efeito Pré-Calculado...**

ou

**Analisar > Metanálise > Metarregressão**

2. No diálogo **Metanálise**, clique em **Criar gráfico** e em seguida, clique na guia **Gráfico de bolhas**.

3. Selecione e defina as configurações de gráfico de bolhas apropriadas.

4. Clique **Continuar**.

## **Gráfico de funil**

A guia **Gráfico de funil** fornece configurações para controlar gráficos de plot de funil que são exibidos na saída para metanálise com resultados contínuos e binários quando dados brutos e dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### **Gráfico de funil**

A configuração opcional ativa e desativa a saída de gráfico de funil. A configuração está desativada por padrão.



### Valores de eixo Y

Controla os valores de eixo Y.

#### Erro padrão

Cria um gráfico do erro padrão.

#### Erro padrão inverso

Cria um gráfico do erro padrão inverso.

#### Variância

Cria gráfico da variação.

#### Variância inversa

Cria gráfico da variação inversa.

### Intervalo de eixo X

Especifica a amplitude do gráfico do eixo X. Quando a configuração é ativada, dois valores numéricos devem ser especificados e atender o  $\text{Limite Superior} > \text{Limite Inferior}$ . Observe que os valores especificados se aplicam a todos os gráficos de funil.

Para metanálise, com resultados binários, quando a configuração **Tamanho do Efeito** é definida para **Diferença de Risco**, ela também deve atender o  $\text{Limite Superior} \leq 1$  e o  $\text{Limite Inferior} \geq -1$ .

### Intervalo de eixo Y

Especifica a amplitude do gráfico do eixo Y. Quando a configuração é ativada, dois valores numéricos devem ser especificados e atender o  $\text{Limite Superior} > \text{Limite Inferior} \geq 0$ . Observe que os valores especificados se aplicam a todos os gráficos de funil.

### Incluir estudos imputados com corte e preenchimento

Controla se os estudos imputados são incluídos no gráfico de funil quando a configuração **Estimar número de estudos ausentes** está ativada na caixa de diálogo **Cortar e preencher**. As configurações estão desabilitadas por padrão, o que exclui os estudos imputados. Quando **Exibir tamanho do efeito geral dos estudos observados** é selecionado, uma linha de referência vertical que indica o tamanho do efeito geral estimado é adicionada. A configuração leva em conta os estudos observados e imputados.

### Variáveis

Lista as variáveis de conjunto de dados disponíveis.

### Rótulo

Opcionalmente, especifique uma variável sobre a qual o gráfico de funil é rotulado. A lista **Colocação** fornece opções de colocação automática de rótulo (**Automático**) ou de colocação **À direita, À esquerda, Na parte superior** e **Na parte inferior**.

## Definindo configurações de gráfico de funil

1. No menu, escolha:

**Analisar > Metanálise > Resultados Contínuos ou Resultados Binários > Dados Brutos... ou Tamanho do Efeito Pré-Calculado...**

2. No diálogo **Metanálise**, clique em **Criar gráfico** e, em seguida, clique na guia **Gráfico de funil**.

3. Selecione e defina as configurações apropriadas de gráfico de funil.

4. Clique **Continuar**.

## Gráfico de Galbraith

A guia **Gráfico de Galbraith** fornece configurações para controlar gráficos de plot de Galbraith, que são exibidos na saída para metanálise, com resultados contínuos e binários quando dados brutos e dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### Gráfico de Galbraith

A configuração opcional ativa e desativa a saída do gráfico de bolhas. Este serviço está desativado por padrão.

### Exibir limites de intervalo de confiança

Controla a exibição de limites de intervalo de confiança. A configuração é ativada por padrão, o que inclui os limites do intervalo de confiança no gráfico de bolhas. Quando a configuração está desativada, os limites do intervalo de confiança não são incluídos no gráfico de bolhas.

### Intervalo de eixo X

Especifica a amplitude do gráfico do eixo X. Quando a configuração é ativada, dois valores numéricos devem ser especificados e atender o  $\text{Limite Superior} > \text{Limite Inferior} \geq 0$ . Observe que os valores especificados se aplicam a todos os gráficos de Galbraith.

### Intervalo de eixo Y

Especifica a amplitude do gráfico do eixo Y. Quando a configuração é ativada, dois valores numéricos devem ser especificados e atender o  $\text{Limite Superior} > \text{Limite Inferior}$ . Observe que os valores especificados se aplicam a todos os gráficos de Galbraith.

### Variáveis

Lista as variáveis de conjunto de dados disponíveis.

### Rótulo

Opcionalmente, especifique uma variável sobre a qual o gráfico de bolhas é rotulado. A lista **Colocação** fornece opções de colocação automática de rótulo (**Automático**) ou de colocação **À direita, À esquerda, Na parte superior e Na parte inferior**.

## Definindo configurações de gráfico de Galbraith

1. A partir dos menus, escolha:

**Analisar > Metanálise > Resultados Contínuos ou Resultados Binários > Dados Brutos... ou Tamanho do Efeito Pré-Calculado...**

2. No diálogo **Metanálise**, clique em **Criar gráfico** e, em seguida, clique na guia **Gráfico de Galbraith**.

3. Selecione e defina as configurações de gráfico de Galbraith apropriadas.

4. Clique **Continuar**.

## Gráfico de L'Abbé

A guia **E plot de L'Abbé** fornece configurações para controlar os gráficos de e plot de L'Abbé, que são exibidos na saída para metanálise, com resultados contínuos e binários quando dados brutos e dados de tamanho do efeito pré-calculados são fornecidos no conjunto de dados ativo.

### E plot de L'Abbé

A configuração opcional ativa e desativa a saída de e plot de L'Abbé. Este serviço está desativado por padrão.

### Desenhar pontos proporcionais às ponderações

Controla como os pontos são desenhados em relação às ponderações. A configuração é ativada por padrão, que desenha os pontos proporcionais aos pesos. Quando a configuração está desativada, todos os pontos são desenhados com o mesmo tamanho.

### Linhas de referência

#### Tamanho geral do efeito

Controla a exibição da linha de referência que representa o tamanho do efeito geral estimado. A configuração é desativada por padrão, o que suprime a linha de referência. A linha de referência é exibida quando a configuração é ativada.

#### Tamanho do efeito nulo

Controla a exibição da linha de referência que representa a ausência de efeitos. A configuração é ativada por padrão, que exibe a linha de referência. A linha de referência é suprimida quando a configuração não é ativada.

### Intervalo de eixo X

Especifica a amplitude do gráfico do eixo X. Quando a configuração é ativada, dois valores numéricos devem ser especificados e atender o  $\text{Limite Superior} > \text{Limite Inferior}$ . Observe que os valores especificados se aplicam a todos os gráficos de Galbraith.

Para metanálise, com resultados binários, quando a configuração **Tamanho do Efeito** é definida para **Diferença de Risco**, ela também deve atender o Limite Inferior  $\geq 0$  e o Limite Superior  $\leq 1$ . Quando a configuração **Tamanho do Efeito** é definida para **Razão de Risco de Log**, ela deve atender o Limite Superior  $\leq 0$ .

### Intervalo de eixo Y

Especifica a amplitude do gráfico do eixo Y. Quando a configuração é ativada, dois valores numéricos devem ser especificados e atender o Limite Superior  $>$  Limite Inferior. Observe que os valores especificados se aplicam a todos os gráficos de Galbraith.

Para metanálise, com resultados binários, quando a configuração **Tamanho do Efeito** é definida para **Diferença de Risco**, ela também deve atender o Limite Inferior  $\geq 0$  e o Limite Superior  $\leq 1$ . Quando a configuração **Tamanho do Efeito** é definida para **Razão de Risco de Log**, ela deve atender o Limite Superior  $\leq 0$ .

### Variáveis

Lista as variáveis de conjunto de dados disponíveis.

### Rótulo

Opcionalmente, especifique uma variável sobre a qual o gráfico de bolhas é rotulado. A lista **Colocação** fornece opções de colocação automática de rótulo (**Automático**) ou de colocação **À direita, À esquerda, Na parte superior** e **Na parte inferior**.

## Definindo configurações de e plot de L'Abbé

1. A partir dos menus, escolha:

**Analisar > Análise de meta > Resultados contínuos ou resultados binários > Dados brutos...**

2. No diálogo **Metanálise**, clique em **Criar gráfico** e em seguida, clique na guia **e plot de L'Abbé**.

3. Selecione e defina as configurações de e plot de L'Abbé apropriadas.

4. Clique **Continuar**.

## Livro de códigos

---

Codebook relata as informações de dicionário - como nomes de variáveis, rótulos de variáveis, rótulos de valor, valores omissos - e estatísticas básicas para todas ou somente as variáveis e conjuntos de múltiplas respostas especificados no conjunto de dados ativo. Para variáveis nominais e ordinais e conjuntos de múltiplas respostas, as estatísticas básicas incluem contagens e porcentagens. Para variáveis de escala, as estatísticas básicas incluem a média, o desvio padrão e quartis.

Nota: o Codebook ignora o status de arquivo dividido. Isto inclui grupos de arquivo dividido criados para imputação múltipla de valores omissos (disponíveis na opção de complemento de Valores Omissos).

Para obter um Codebook

1. A partir dos menus, escolha:

**Analisar > Relatórios > Codebook**

2. Clique na guia Variáveis.

3. Selecione uma ou mais variáveis e/ou conjuntos de múltiplas respostas.

Opcionalmente, é possível:

- Controlar as informações de variável que são exibidas.
- Controlar as estatísticas que são exibidas (ou excluir todas as estatísticas básicas).
- Controlar a ordem na qual as variáveis e os conjuntos de múltiplas respostas são exibidos.
- Mudar o nível de medição para qualquer variável na lista de origem para mudar as estatísticas básicas exibidas. Consulte o tópico [“Guia Estatísticas do livro de códigos” na página 86](#) para obter mais informações

Mudando o nível de medição

É possível mudar temporariamente o nível de medição para variáveis. (Não é possível mudar o nível de medição para conjuntos de múltiplas respostas. Eles são sempre tratados como nominais).

1. Clique com o botão direito em uma variável na lista de origem.
2. Selecione um nível de medição no menu pop-up.

Isso altera o nível de medição temporariamente. Em termos práticos, isso é útil apenas para variáveis numéricas. O nível de medição para variáveis de sequência de caracteres é restrito a nominal ou ordinal, que são tratadas da mesma forma pelo procedimento Codebook.

## Guia Saída do livro de códigos

A guia Saída controla as informações de variáveis incluídas para cada variável e conjunto de múltiplas respostas, a ordem em que as variáveis e conjuntos de múltiplas respostas são exibidos e o conteúdo da tabela de informações de arquivo opcional.

Informações da variável

Isso controla as informações do dicionário exibidas para cada variável.

**Posição.** Um inteiro que representa a posição da variável na ordem do arquivo. Isso não está disponível para conjuntos de múltiplas respostas.

**Rótulo.** O rótulo descritivo associado à variável ou conjunto de múltiplas respostas.

**Tipo.** Tipo de dados fundamentais. Este é *Númérico*, *Sequência de Caracteres* ou *Conjunto de Múltiplas Respostas*.

**Formato.** O formato de exibição para a variável, como *A4*, *F8.2* ou *DATE11*. Isso não está disponível para conjuntos de múltiplas respostas.

**Nível de medição.** Os valores possíveis são *Nominal*, *Ordinal*, *Escala* e *Desconhecido*. O valor exibido é o nível de medição armazenado no dicionário e não é afetado por nenhuma substituição temporária de nível de medição especificada pela alteração do nível de medição na lista de variáveis de origem na guia Variáveis. Isso não está disponível para conjuntos de múltiplas respostas.

Nota: o nível de medição para variáveis numéricas pode ser "desconhecido" antes da primeira passagem de dados quando o nível de medição não foi definido explicitamente, como dados lidos de uma fonte externa ou variáveis recém-criadas. Veja o tópico para obter mais informações.

**Papel.** Algumas caixas de diálogo suportam a capacidade de pré-selecionar variáveis para análise com base em funções definidas. para obter mais informações.

**Rótulo de valor.** Rótulos descritivos associados a valores de dados específicos. para obter mais informações.

- Se Contagem ou Porcentagem for selecionado na guia Estatísticas, os rótulos de valor definidos serão incluídos na saída, mesmo que você não selecione os rótulos de valor aqui.
- Para conjuntos de múltiplas dicotomias, "rótulos de valor" são os rótulos de variáveis para as variáveis elementares no conjunto ou os rótulos de valores contados, dependendo de como o conjunto é definido. Veja o tópico para obter mais informações.

**Valores omissos.** Valores omissos definidos pelo usuário. para obter mais informações. Se Contagem ou Porcentagem for selecionado na guia Estatísticas, os rótulos de valor definidos serão incluídos na saída, mesmo que você não selecione Valores ausentes aqui. Isso não está disponível para conjuntos de múltiplas respostas.

**Atributos customizados.** Atributos de variáveis personalizadas definidas pelo usuário. A saída inclui os nomes e os valores de quaisquer atributos de variável customizados associados a cada variável. Consulte o tópico para obter mais informações. Isso não está disponível para conjuntos de múltiplas respostas.

**Atributos reservados.** Atributos de variável de sistema reservados. É possível exibir os atributos do sistema, mas você não deve mudá-los. Os nomes dos atributos do sistema começam com um cifrão (\$). Atributos que não são de exibição, com nomes que começam com "@" ou "\$@", não são incluídos. A saída

inclui os nomes e valores para quaisquer atributos do sistema associados a cada variável. Isso não está disponível para conjuntos de múltiplas respostas.

#### Informações do arquivo

A tabela de informações de arquivo opcional pode incluir qualquer um dos seguintes atributos de arquivo:

**Nome do arquivo.** Nome do arquivo de dados do IBM SPSS Statistics. Se o conjunto de dados nunca tiver sido salvo no formato IBM SPSS Statistics, então não há nenhum nome do arquivo de dados. (Se não houver nenhum nome de arquivo exibido na barra de título da janela Editor de Dados, o conjunto de dados ativo não terá um nome de arquivo.)

**Localização.** Localização do diretório (pasta) do arquivo de dados IBM SPSS Statistics. Se o conjunto de dados nunca tiver sido salvo no formato IBM SPSS Statistics, então não há nenhuma localização.

**Número de casos.** Número de casos no conjunto de dados ativo. Este é o número total de casos, incluindo todos os casos que podem ser excluídos das estatísticas de resumo devido às condições do filtro.

**Rótulo.** Esse é o rótulo de arquivo (se houver) definido pelo comando `FILE LABEL`.

**Documentos.** Texto do documento do arquivo de dados. para obter mais informações.

**Status de ponderação.** Se a ponderação estiver ativada, o nome da variável de ponderação será exibido. Veja o tópico para obter mais informações.

**Atributos customizados.** Atributos de arquivo de dados customizados definidos pelo usuário. Atributos de arquivo de dados definidos com o comando `DATAFILE ATTRIBUTE`.

**Atributos reservados.** Atributos de arquivo de dados do sistema reservados. É possível exibir os atributos do sistema, mas você não deve mudá-los. Os nomes dos atributos do sistema começam com um cifrão (\$) . Atributos que não são de exibição, com nomes que começam com "@" ou "\$@", não são incluídos. A saída inclui os nomes e valores para quaisquer atributos de arquivo de dados do sistema.

#### Ordem de exibição da variável

As alternativas a seguir estão disponíveis para controlar a ordem na qual as variáveis e os conjuntos de múltiplas respostas são exibidos.

**Alfabética.** Ordem alfabética por nome de variável.

**Arquivo.** A ordem em que as variáveis aparecem no conjunto de dados (a ordem em que são exibidas no Editor de Dados). Em ordem crescente, os conjuntos de múltiplas respostas são exibidos por último, depois de todas as variáveis selecionadas.

**Nível de medição.** Ordenar por nível de medição. Isso cria quatro grupos de classificação: nominal, ordinal, escala e desconhecido. Conjuntos de múltiplas respostas são tratados como nominais.

Nota: o nível de medição para variáveis numéricas pode ser "desconhecido" antes da primeira passagem de dados quando o nível de medição não foi definido explicitamente, como dados lidos de uma fonte externa ou variáveis recém-criadas.

**Lista de variáveis.** A ordem em que as variáveis e os conjuntos de múltiplas respostas aparecem na lista de variáveis selecionadas na guia Variáveis.

**Nome do atributo customizado.** A lista de opções de ordem de classificação também inclui os nomes de quaisquer atributos de variáveis customizadas definidas pelo usuário. Em ordem crescente, as variáveis que não possuem o atributo são classificadas para o topo, seguidas pelas variáveis que possuem o atributo mas sem valor definido para o atributo, seguidas por variáveis com valores definidos para o atributo em ordem alfabética dos valores.

#### Número máximo de categorias

Se a saída incluir rótulos de valor, contagens ou porcentagens para cada valor exclusivo, você poderá suprimir essas informações da tabela se o número de valores exceder o valor especificado. Por padrão, essas informações são suprimidas se o número de valores exclusivos da variável exceder 200.

## Guia Estatísticas do livro de códigos

A guia Estatísticas permite controlar as estatísticas básicas incluídas na saída ou suprimir totalmente a exibição das estatísticas básicas.

### Contagens e porcentagens

Para variáveis nominais e ordinais, conjuntos de múltiplas respostas e valores rotulados de variáveis de escala, as estatísticas disponíveis são:

*Contagem.* A contagem ou o número de casos que possuem cada valor (ou intervalo de valores) de uma variável.

*Porcentagem.* A porcentagem de casos que possuem um valor específico.

### Dispersão e tendência central

Para variáveis de escala, as estatísticas disponíveis são:

*Média.* Uma medida de tendência central A média aritmética, a soma dividida pelo número de casos.

*desvio padrão.* Medida de dispersão em torno da média. Em uma distribuição normal, 68% dos casos estão dentro de um desvio padrão da média e 95% dos casos estão dentro de dois desvios padrão. Por exemplo, se a idade média for 45, com um desvio padrão de 10, 95% dos casos estariam entre 25 e 65 em uma distribuição normal.

*Quartis.* Exibe os valores correspondentes aos 25°, 50° e 75° percentis.

Nota: é possível mudar temporariamente o nível de medição associado a uma variável (e, assim, alterar as estatísticas básicas exibidas para essa variável) na lista de variáveis de origem na guia Variáveis.

## Frequências

O procedimento Frequências fornece exibições estatísticas e gráficas que são úteis para descrever muitos tipos de variáveis. O procedimento Frequências é um bom lugar para começar a observar seus dados.

Para um relatório de frequência e gráfico de barras, é possível organizar os valores distintos em ordem crescente ou decrescente ou pode ordenar as categorias por suas frequências. O relatório de frequências pode ser suprimido quando uma variável tem muitos valores distintos. É possível rotular gráficos com frequências (o padrão) ou porcentagens.

### Exemplo

Qual é a distribuição de clientes de uma empresa por tipo de mercado? A partir do resultado, é possível descobrir que 37,5% de seus clientes estão em agências governamentais, 24,9% estão em corporações, 28,1% estão em instituições acadêmicas e 9,4% estão no setor de saúde. Para dados quantitativos contínuos, como receita de vendas, você pode descobrir que a venda média do produto é de US\$ 3.576 com um desvio padrão de US\$ 1.078.

### Estatísticas e gráficos

Contagens de frequência, porcentagens, porcentagens cumulativas, média, mediana, moda, soma, desvio padrão, variância, intervalo, valores mínimo e máximo, erro padrão da média, assimetria e curtose (ambos com erros padrão), quartis, percentis especificados pelo usuário, gráficos de barras, gráficos de pizza e histogramas.

## Considerações de dados

### Dados

Utilize códigos ou sequências de caracteres numéricos para codificar variáveis categóricas (níveis de medição nominais ou ordinais).

### Suposições

As tabulações e porcentagens fornecem uma descrição útil para dados de qualquer distribuição, especialmente para variáveis com categorias ordenadas ou não ordenadas. A maioria das estatísticas de resumo opcionais, como a média e o desvio padrão, são baseadas na teoria normal e são apropriadas para variáveis quantitativas com distribuições simétricas. Estatísticas robustas, como

mediana, quartis e percentis, são apropriadas para variáveis quantitativas que podem ou não atender à suposição de normalidade.

## Obtendo tabelas de frequências

1. A partir dos menus, escolha:

**Analisar > Estatísticas Descritivas > Frequências...**

2. Selecione uma ou mais variáveis categóricas ou quantitativas.

3. Opcionalmente, selecione **Criar tabelas de estilo APA** para criar tabelas de saída que seguem as diretrizes de estilo APA.

4. Opcionalmente, é possível:

- Clicar em **Estatísticas** para estatísticas descritivas para variáveis quantitativas.
- Clicar em **Gráficos** para gráficos de barras, gráficos de setores circulares e histogramas.
- Clicar em **Formatar** para a ordem na qual os resultados são exibidos.
- Clique em **Estilo** para especificar condições para mudar automaticamente as propriedades de tabelas dinâmicas com base em condições específicas.
- Clique em **Autoinicialização** para derivar estimativas robustas de erros padrão e intervalos de confiança para estimativas, como a média, mediana, proporção, razão de chances, coeficiente de correlação ou coeficiente de regressão. Ela também pode ser usada para criar testes de hipótese.

## Estatísticas de frequências

**Valores de Percentil.** Valores de uma variável quantitativa que dividem os dados ordenados em grupos de forma que uma determinada porcentagem fique acima e outra abaixo. Os quartis (os percentis 25, 50 e 75) dividem as observações em quatro grupos de igual tamanho. Se desejar um número igual de grupos diferente de quatro, selecione **Pontos de corte para n grupos iguais**. Você também pode especificar percentis individuais (por exemplo, o percentil 95, o valor abaixo do qual 95% das observações caem).

**Tendência Central.** As estatísticas que descrevem a localização da distribuição incluem a média, a mediana, a moda e a soma de todos os valores.

- *Média.* Uma medida de tendência central A média aritmética, a soma dividida pelo número de casos.
- *Median.* O valor acima e abaixo do qual metade dos casos cai, o 50th percentil. Se houver um número par de casos, a mediana é a média dos dois casos intermediários quando estão classificados em ordem crescente ou decrescente. A mediana é uma medida da tendência central não sensível a valores excessivos (ao contrário da média, que pode ser afetada por alguns valores extremamente altos ou baixos).
- *Modo.* O valor que ocorre com mais frequência. Se vários valores compartilharem a maior frequência da ocorrência, cada um deles será um modo. O procedimento Frequências relata apenas o menor desses diversos modos.
- *Sum.* A soma ou o total dos valores, em todos os casos com valores não omissos.

**Dispersão.** As estatísticas que medem a quantidade de variação ou dispersão nos dados incluem o desvio padrão, variância, intervalo, mínimo, máximo e erro padrão da média.

- *Desvio padrão.* Medida de dispersão em torno da média. Em uma distribuição normal, 68% dos casos estão dentro de um desvio padrão da média e 95% dos casos estão dentro de dois desvios padrão. Por exemplo, se a idade média for 45, com um desvio padrão de 10, 95% dos casos estariam entre 25 e 65 em uma distribuição normal.
- *Variância.* Uma medida de dispersão ao redor da média, igual à soma dos desvios quadrados da média dividido por um menor que o número de casos. A variância é medida em unidades que são o quadrado daquelas da própria variável.
- *Intervalo.* A diferença entre os valores maior e menor de uma variável numérica, o máximo menos o mínimo.

- *Mínimo*. O menor valor de uma variável numérica.
- *Máximo*. O maior valor de uma variável numérica.
- *S. E. média*. Uma medida de quanto o valor da média pode variar de amostra para amostra obtida da mesma distribuição. Ela pode ser utilizada para comparar aproximadamente a média observada com um valor hipotético (ou seja, é possível concluir que os dois valores serão diferentes se a razão da diferença com o erro padrão for inferior a -2 ou superior a +2).

**Distribuição.** A assimetria e a curtose são estatísticas que descrevem a forma e a simetria da distribuição. Essas estatísticas são exibidas com seus erros padrão.

- *Assimetria*. Uma medida da assimetria de uma distribuição. A distribuição normal é simétrica e tem um valor de assimetria de 0. Uma distribuição com assimetria positiva significativa tem um longo rodapé direito. Uma distribuição com assimetria negativa significativa tem um longo rodapé esquerdo. Como uma orientação, um valor de assimetria mais de duas vezes seu erro padrão é obtido para indicar uma partida de simetria.
- *Curtose*. Uma medida da extensão até a qual há valores discrepantes. Para a uma distribuição normal, o valor da estatística da curtose é zero. A curtose positiva indica que os dados exibem mais valores discrepantes que uma distribuição normal. A curtose negativa indica que os dados exibem menos valores discrepantes extremos do que uma distribuição normal. A definição de curtose que é usada, em que o valor é de 0 para uma distribuição normal, é, às vezes, referida como excesso de curtose. Alguns softwares podem reportar curtose tal que o valor é 3 para uma distribuição normal.

**Os valores são pontos médios do grupo.** Se os valores em seus dados forem pontos médios de grupos (por exemplo, as idades de todas as pessoas na casa dos trinta são codificadas como 35), selecione esta opção para estimar a mediana e os percentis dos dados originais não agrupados.

## Gráficos de frequências

**Nota:** Os gráficos não são produzidos na saída quando **Executar autoinicialização** está ativado no diálogo Autoinicialização.

### Chart Type

Um gráfico de pizza exibe a contribuição de partes para um todo. Cada fatia de um gráfico de pizza corresponde a um grupo definido por uma única variável de agrupamento. Um gráfico de barras exibe a contagem para cada valor ou categoria distinta como uma barra separada, permitindo comparar as categorias visualmente. Um histograma também possui barras, mas eles são representados ao longo de uma escala de intervalo igual. A altura de cada barra é a contagem de valores de uma variável quantitativa dentro do intervalo. Um histograma mostra a forma, o centro e a dispersão da distribuição. Uma curva normal sobreposta a um histograma ajuda a avaliar se os dados são distribuídos normalmente.

### Valores do gráfico

Para gráficos de barras, o eixo de escala pode ser rotulado por contagens ou porcentagens de frequência.

## Formato de frequências

**Ordenar por.** A tabela de frequência pode ser organizada de acordo com os valores reais nos dados ou de acordo com a contagem (frequência de ocorrência) desses valores, e a tabela pode ser organizada em ordem crescente ou decrescente. No entanto, se você solicitar um histograma ou percentis, o Frequencies assume que a variável é quantitativa e exibe seus valores em ordem crescente.

**Diversas Variáveis.** Se você produzir tabelas de estatísticas para variáveis múltiplas, será possível exibir todas as variáveis em uma única tabela (**Comparar variáveis**) ou exibir uma tabela de estatísticas separada para cada variável (**Organizar saída por variáveis**).

**Suprimir tabelas com muitas categorias.** Esta opção impede a exibição de tabelas com mais do que o número especificado de valores.



## Descritivas

---

O procedimento Descritivos exibe estatísticas básicas univariadas de diversas variáveis em uma única tabela e calcula valores padronizados (escores  $z$ ). As variáveis podem ser classificadas pelo tamanho de suas médias (em ordem crescente ou decrescente), em ordem alfabética ou pela ordem na qual você seleciona as variáveis (o padrão).

Quando escores  $z$  são salvos, eles são incluídos nos dados no Editor de Dados e estão disponíveis para gráficos, listagens de dados e análises. Quando as variáveis são gravadas em unidades diferentes (por exemplo, produto interno bruto per capita e em porcentagem), uma transformação de escore  $z$  colocará as variáveis em uma escala comum para facilitar a comparação visual.

**Exemplo.** Se cada caso em seus dados contiver os totais diários de vendas para cada membro da equipe de vendas (por exemplo, uma entrada para Bob, uma entrada para Kim e uma entrada para Brian) coletados diariamente por vários meses, o procedimento Descritivos pode calcular a média diária de vendas para cada membro da equipe e pode ordenar os resultados da média de vendas mais alta para a média de vendas mais baixa.

**Estatísticas.** Tamanho da amostra, média, mínimo, máximo, desvio padrão, variância, intervalo, soma, erro padrão da média e curtose e assimetria com seus erros padrão.

Considerações sobre dados descritivos

**Dados.** Utilize variáveis numéricas após verificá-las graficamente para registrar erros, valores discrepantes e anomalias de distribuição. O procedimento Descritivos é muito eficiente para arquivos grandes (milhares de casos).

**Suposições.** A maioria das estatísticas disponíveis (incluindo escores  $z$ ) baseia-se em teoria normal e é apropriada para variáveis quantitativas (medidas de nível de intervalo ou de razão) com distribuições simétricas. Evite variáveis com categorias não ordenadas ou distribuições distorcidas. A distribuição de escores  $z$  possui a mesma forma que a dos dados originais, portanto, calcular escores  $z$  não é uma correção para dados de problema.

Para obter estatísticas descritivas

1. A partir dos menus, escolha:

**Analisar > Estatísticas Descritivas > Descritivos...**

2. Selecione duas ou mais variáveis.

Opcionalmente, é possível:

- Selecionar **Salvar valores padronizados como variáveis** para salvar escores  $z$  como novas variáveis.
- Clicar em **Opções** para estatísticas opcionais e a ordem de exibição.

## Opções Descritivas

**Média e Soma.** A média, ou média aritmética, é exibida por padrão.

**Dispersão.** Estatísticas que medem a dispersão ou a variação nos dados incluem o desvio padrão, variância, intervalo, mínimo, máximo e o erro padrão da média.

- *Desvio padrão.* Medida de dispersão em torno da média. Em uma distribuição normal, 68% dos casos estão dentro de um desvio padrão da média e 95% dos casos estão dentro de dois desvios padrão. Por exemplo, se a idade média for 45, com um desvio padrão de 10, 95% dos casos estariam entre 25 e 65 em uma distribuição normal.
- *Variância.* Uma medida de dispersão ao redor da média, igual à soma dos desvios quadrados da média dividido por um menor que o número de casos. A variância é medida em unidades que são o quadrado daquelas da própria variável.
- *Intervalo.* A diferença entre os valores maior e menor de uma variável numérica, o máximo menos o mínimo.
- *Mínimo.* O menor valor de uma variável numérica.

- *Máximo*. O maior valor de uma variável numérica.
- *Média S.E.* Uma medida de quanto o valor da média pode variar de amostra para amostra obtida da mesma distribuição. Ela pode ser utilizada para comparar aproximadamente a média observada com um valor hipotético (ou seja, é possível concluir que os dois valores serão diferentes se a razão da diferença com o erro padrão for inferior a -2 ou superior a +2).

**Distribuição.** Curtose e assimetria são estatísticas que caracterizam a forma e a simetria da distribuição. Essas estatísticas são exibidas com seus erros padrão.

- *Curtose*. Uma medida da extensão até a qual há valores discrepantes Para a uma distribuição normal, o valor da estatística da curtose é zero. A curtose positiva indica que os dados exibem mais valores discrepantes que uma distribuição normal. A curtose negativa indica que os dados exibem menos valores discrepantes extremos do que uma distribuição normal. A definição de curtose que é usada, em que o valor é de 0 para uma distribuição normal, é, às vezes, referida como excesso de curtose. Alguns softwares podem reportar curtose tal que o valor é 3 para uma distribuição normal.
- *Assimetria*. Uma medida da assimetria de uma distribuição A distribuição normal é simétrica e tem um valor de assimetria de 0. Uma distribuição com assimetria positiva significativa tem um longo rodapé direito. Uma distribuição com assimetria negativa significativa tem um longo rodapé esquerdo. Como uma orientação, um valor de assimetria mais de duas vezes seu erro padrão é obtido para indicar uma partida de simetria.

**Ordem de Exibição.** Por padrão, as variáveis são exibidas na ordem em que você as selecionou. Opcionalmente, é possível exibir variáveis em ordem alfabética, por meios crescentes ou decrescentes.

## Recursos adicionais do comando DESCRIPTIVES

O idioma da sintaxe de comando também permite:

- Salvar escores padronizados (escores z) para algumas, mas não todas as variáveis (com o subcomando VARIABLES).
- Especificar nomes para novas variáveis que contêm os escores padronizados (com o subcomando VARIABLES).
- Excluir da análise casos com valores omissos para qualquer variável (com o subcomando MISSING).
- Ordenar as variáveis na exibição pelo valor de quaisquer estatísticas, não apenas pela média (com o subcomando SORT).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Percentis

---

O procedimento Percentis exibe tabelas de percentis..

### Valores de percentis

Ao selecionar **Valores de Percentil**, por padrão, exibe os valores para os 5th, 10th, 25th, 50th, 75th, 90th e 95th percentis. Valores de uma variável quantitativa que dividem os dados ordenados em grupos de modo que uma determinada porcentagem esteja acima e outra porcentagem esteja abaixo.

### Quartis

Quartis (os 25º, 50º e 75º percentis) dividem as observações em quatro grupos de tamanhos iguais.

### Customizado

Ao selecionar **Customizado**, insira pelo menos um valor para executar a análise. Os valores de entrada devem ser um número no intervalo de 0 a 100. Use os botões **Incluir**, **Alterar** **Remover** para trabalhar com os valores na lista de valores de percentil.

### Método de percentil

Por padrão, o método HAVERAGE é selecionado para calcular percentis.

## Obtendo tabelas de percentil

1. Nos menus, escolha:

**Analisar > Estatísticas Descritivas > Percentis ..**

2. Selecione uma ou mais variáveis numéricas e mova-as para o campo Variáveis.

3. Selecione **Valores de percentil** para especificar valores para calcular o percentil.

4. Selecione **Método Percentil** para selecionar o método de cálculo de percentual.

5. Como opção, você pode:

- Clique em **Ausente**, para controlar o tratamento de valores omissos.
- Clique em **Bootstrap** para derivar estimativas robustas de erros padrão e intervalos de confiança para estimativas de percentuais.

## Valores omissos de percentis

### Valores omissos

Controla o tratamento de valores omissos.

### Excluir casos de método listwise

Os casos com valores omissos para qualquer variável especificada são excluídos de todas as análises. Este é o padrão.

### Por variável

Excluir casos com valores omissos em uma base variável por variável.

## Explorar

---

O procedimento Explorar produz estatísticas básicas e exibições gráficas, tanto para todos os casos ou separadamente para grupos de casos. Há muitos motivos para utilizar o procedimento Explorar - triagem de dados, identificação de valor discrepante, descrição, verificação de suposição e caracterização das diferenças entre as subpopulações (grupos de casos). A triagem de dados pode mostrar que você tem valores incomuns, valores extremos, lacunas nos dados ou outras peculiaridades. Explorar os dados pode ajudar a determinar se as técnicas estatísticas que você está considerando para análise de dados são apropriadas. A exploração pode indicar que você precisa transformar os dados se a técnica exigir uma distribuição normal. Ou você pode decidir que precisa de testes não paramétricos.

**Exemplo.** Observe a distribuição dos tempos de aprendizado do labirinto para ratos sob quatro esquemas de reforço diferentes. Para cada um dos quatro grupos, é possível ver se a distribuição dos tempos é aproximadamente normal e se as quatro variações são iguais. Também é possível identificar os casos com os cinco maiores e os cinco menores tempos. Os boxplots e gráficos de ramos e folhas resumem graficamente a distribuição dos tempos de aprendizado para cada um dos grupos.

**Estatísticas e gráficos.** Média, mediana, 5% da média aparada, erro padrão, variância, desvio padrão, mínimo, máximo, intervalo, amplitude interquartil, assimetria e curtose e seus erros padrão, intervalo de confiança para a média (e o nível de confiança especificado), percentis, estimador M de Huber, estimador de onda de Andrews, estimador M redescendente de Hampel, estimador de bponderação de Tukey, os cinco maiores e os cinco menores valores, estatística de Kolmogorov-Smirnov com um nível de significância de Lilliefors para teste de normalidade, e a estatística de Shapiro-Wilk. Boxplots, gráficos de ramos e folhas, histogramas, gráficos de normalidade e gráficos de dispersão versus nível com testes e transformações de Levene.

Considerações de Explorar Dados

**Dados.** O procedimento Explorar pode ser utilizado para variáveis quantitativas (medidas de nível de intervalo ou razão). Uma variável de fator (utilizada para dividir os dados em grupos de casos) deve ter um número razoável de valores distintos (categorias). Estes valores podem ser sequência curta ou numéricos. A variável de rótulo case, utilizada para rotular valores discrepantes em boxplots, pode ser uma sequência curta, uma sequência longa (primeiros 15 bytes) ou numérica.

**Suposições.** A distribuição dos dados não precisa ser simétrica ou normal.

Para Explorar Seus Dados

1. A partir dos menus, escolha:

**Analisar > Estatísticas Descritivas > Explorar...**

2. Selecione uma ou mais variáveis dependentes.

Opcionalmente, é possível:

- Selecione uma ou mais variáveis de fator, cujos valores definirão grupos de casos.
- Selecione uma variável de identificação para rotular os casos.
- Clicar em **Estatísticas** para estimadores robustos, valores discrepantes, percentis e tabelas de frequência.
- Clicar em **Gráficos** para histogramas, gráficos e testes de probabilidade normal e gráficos de dispersão versus nível com estatísticas de Levene.
- Clicar em **Opções** para o tratamento de valores omissos.

## Explorar Estatísticas

**Descritivos.** Essas medidas de tendência central e de dispersão são exibidas por padrão. As medidas de tendência central indicam a localização da distribuição; elas incluem a média, a mediana e 5% da média aparada. As medidas de dispersão mostram a dissimilaridade dos valores; essas incluem erro padrão, variância, desvio padrão, mínimo, máximo, intervalo e a amplitude interquartil. As estatísticas descritivas também incluem medidas da forma da distribuição; assimetria e curtose são exibidas com seus erros padrão. O intervalo de confiança de nível de 95% para a média também é exibido; é possível especificar um nível de confiança diferente.

**Estimadores M.** Alternativas robustas para a média e mediana da amostra para estimar a localização. Os estimadores que são calculados diferem nos pesos que eles aplicam aos casos. Estimador M de Huber, o estimador de onda de Andrews, o estimador-M redescendente de Hampel e o estimador de bponderação de Tukey.

**Valores Discrepantes.** Exibe os cinco maiores e os cinco menores valores com rótulos case.

### Percentis

O procedimento Percentis exibe tabelas de percentis.. Ao selecionar **Percentil**, por padrão, exibe os valores para os percentis 5th, 10th, 25th, 50th, 75th, 90th e 95th . Valores de uma variável quantitativa que dividem os dados ordenados em grupos de modo que uma determinada porcentagem esteja acima e outra porcentagem esteja abaixo.

### Quartis

Quartis (os 25º, 50º e 75º percentis) dividem as observações em quatro grupos de tamanhos iguais.

### Customizado

Ao selecionar **Customizado**, insira pelo menos um valor para executar a análise. Os valores de entrada devem ser um número no intervalo de 0 a 100. Use os botões **Incluir**, **Alterar** **Remover** para trabalhar com os valores na lista de valores de percentil.

### Método

Por padrão, o método HAVERAGE é selecionado para calcular percentis.

## Explorar Gráficos

**Boxplots.** Essas alternativas controlam a exibição de boxplots quando tiver mais de uma variável dependente. **Níveis do fator juntos** gera uma exibição separada para cada variável dependente. Em uma exibição, boxplots são mostrados para cada um dos grupos definidos por uma variável de fator. **Dependentes juntos** gera uma exibição separada para cada grupo definido por uma variável de fator. Em uma exibição, boxplots são mostrados lado a lado para cada uma das variáveis dependentes. Esta

exibição é particularmente útil quando as diferentes variáveis representam uma única característica medida em momentos diferentes.

**Descritivo.** O grupo Descritivo permite escolher gráficos de ramos e folhas e histogramas.

**Gráficos de normalidade com testes.** Exibe gráficos de probabilidade normal e de probabilidade normal de tendência. A estatística de Kolmogorov-Smirnov, com um nível de significância de Lilliefors para teste de normalidade, é exibida. Se ponderações de não número inteiro forem especificadas, a estatística de Shapiro-Wilk será calculada quando o tamanho da amostra ponderada estiver entre 3 e 50. Para nenhuma ponderação ou ponderações de número inteiro, a estatística é calculada quando o tamanho da amostra ponderada estiver entre 3 e 5.000.

**Dispersão vs. Nível com Teste de Levene.** Controla a transformação de dados para gráficos de dispersão versus nível. Para todos os gráficos de dispersão versus nível, a inclinação da linha de regressão e os testes robustos de Levene para a homogeneidade da variância são exibidos. Se selecionar uma transformação, testes de Levene terão como base os dados transformados. Se nenhuma variável de fator for selecionada, os gráficos de dispersão versus nível não serão produzidos. **Estimação de potência** produz um gráfico dos logs naturais das amplitudes interquartis com relação aos logs naturais das medianas para todas as células, bem como uma estimativa da transformação de potência para alcançar variâncias iguais nas células. Um gráfico de dispersão versus nível ajuda a determinar a potência de uma transformação para estabilizar (tornar mais igual) variâncias entre os grupos. **Transformado** permite selecionar uma das alternativas de potência, talvez seguindo a recomendação da estimativa de potência, e produz gráficos de dados transformados. A amplitude interquartil e a mediana dos dados transformados são representados. **Não transformado** produz gráficos de dados brutos. Isso é equivalente a uma transformação com uma potência de 1.

## Explorar Transformações de Potência

Estas são as transformações de poder para gráficos de dispersão versus nível. Para transformar dados, deve-se selecionar uma potência para a transformação. É possível escolher uma das alternativas a seguir:

- **Logarítmica natural.** Transformação por log natural. Este é o padrão.
- **1/raiz quadrada.** Para cada valor de dados, o recíproco da raiz quadrada é calculado.
- **Recíproco.** O recíproco de cada valor de dados é calculado.
- **Raiz quadrada.** A raiz quadrada de cada valor de dados é calculada.
- **Quadrado.** Cada valor de dados é elevado ao quadrado.
- **Cubo.** Cada valor de dados é elevado ao cubo.

## Opções de exploração

**Valores omissos.** Controla o tratamento de valores omissos.

- **Excluir listwise dos casos.** Casos com valores omissos para qualquer variável dependente ou de fator são excluídos de todas as análises. Este é o padrão.
- **Excluir casos entre pares.** Casos sem valores ausentes para variáveis em um grupo (célula) são incluídos na análise desse grupo. O caso pode ter valores ausentes para variáveis usadas em outros grupos.
- **Valores de relatório.** Os valores ausentes para variáveis de fator são tratados como uma categoria separada. Toda a saída é produzida para esta categoria adicional. As tabelas de frequência incluem categorias de valores omissos. Os valores ausentes para uma variável de fator são incluídos, mas rotulados como ausentes.

## Recursos adicionais do comando EXAMINE

O procedimento Explorar utiliza a sintaxe de comando do EXAMINE. O idioma da sintaxe de comando também permite:

- Solicitar saída total e gráficos, além de saída e gráficos para grupos definidos pelas variáveis de fator (com o subcomando TOTAL).

- Especificar uma escala comum para um grupo de boxplots (com o subcomando SCALE).
- Especificar as interações de variáveis de fator (com o subcomando VARIABLES).
- Especificar percentis diferentes dos padrões (com o subcomando PERCENTILES).
- Calcular percentis de acordo com qualquer um dos cinco métodos (com o subcomando PERCENTILES).
- Especificar qualquer transformação de potência para gráficos de dispersão versus nível (com o subcomando PLOT).
- Especificar o número de valores extremos a serem exibidos (com o subcomando STATISTICS).
- Especificar parâmetros para os estimadores M e estimadores robustos de localização (com o subcomando MESTIMATORS).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Tabulações cruzadas

O procedimento Crosstabs forma tabelas de dois fatores e de diversos fatores e fornece uma variedade de testes e medidas de associação para tabelas de dois fatores. A estrutura da tabela e se as categorias são ordenadas determinam qual teste ou medida utilizar.

Com exceção dos coeficientes gama parciais, as estatísticas e as medidas de associação de Tabulações Cruzadas são calculadas separadamente para cada tabela bidirecional. Se você especificar uma linha, uma coluna e um fator de camada (variável de controle), o procedimento Tabulações Cruzadas formará um painel de estatísticas e medidas associadas para cada valor do fator de camada (ou uma combinação de valores para duas ou mais variáveis de controle). Por exemplo, se *gender* for um fator de estrato para uma tabela de *married* (sim, não) com relação a *life* (se a vida é excitante, rotineira ou monótona), os resultados de uma tabela de dois fatores para mulheres serão calculados separadamente dos resultados dos homens e impressos como painéis um seguido do outro.

**Exemplo.** Os clientes de pequenas empresas são mais propensos a serem lucrativos em vendas de serviços (por exemplo, treinamento e consultoria) do que os de empresas maiores? A partir de uma tabulação cruzada, você pode descobrir que a maioria das pequenas empresas (menos de 500 funcionários) gera altos lucros com serviços, enquanto a maioria das grandes empresas (mais de 2.500 funcionários) gera baixos lucros com serviços.

**Estatísticas e medidas de associação.** Qui-quadrado de Pearson, qui-quadrado de razão de verossimilhança, teste de associação linear por linear, teste exato de Fisher, qui-quadrado corrigido de Yates, *r* de Pearson, *r* de Spearman, coeficiente de contingência, *fi*, *V* de Cramér, lambdas simétrica e assimétrica, tau de Goodman e Kruskal, coeficiente de incerteza, gama, *d* de Somers, tau-*b* e tau-*c* de Kendall, coeficiente eta, kappa de Cohen, estimativa de risco relativa, razão de chances, teste de McNemar, estatísticas de Cochran e Mantel-Haenszel e estatísticas de proporções da coluna.

Considerações sobre dados de tabulações cruzadas

**Dados.** Para definir as categorias de cada variável de tabela, use valores de uma variável numérica ou de sequência (oito ou menos bytes). Por exemplo, para *gender*, é possível codificar os dados como 1 e 2 ou como *male* e *female*.

**Suposições.** Algumas estatísticas e medidas assumem categorias ordenadas (dados ordinais) ou valores quantitativos (dados de intervalo ou razão), conforme discutido na seção sobre estatística. Outras são válidas quando as variáveis da tabela possuem categorias não ordenadas (dados nominais). Para as estatísticas baseadas em qui-quadrado (*fi*, *V* de Cramér e o coeficiente de contingência), os dados devem ser uma amostra aleatória de uma distribuição multinomial.

*Nota:* As variáveis ordinais podem ser códigos numéricos que representam categorias (por exemplo, 1 = *low*, 2 = *medium*, 3 = *high*) ou valores de sequência de caracteres. No entanto, supõe-se que a ordem alfabética dos valores da sequência de caracteres reflita a verdadeira ordem das categorias. Por exemplo, para uma variável de sequência de caracteres com os valores *low*, *medium*, *high*, a ordem das categorias é interpretada como *high*, *low*, *medium* - que não é a ordem correta. Em geral, é mais confiável utilizar códigos numéricos para representar dados ordinais.

Para Obter Tabulações Cruzadas

1. No menu, escolha:

**Analisar > Estatísticas Descritivas > Crosstabs...**

2. Selecione uma ou mais variáveis de linha e uma ou mais variáveis de coluna.

Opcionalmente, é possível:

- Selecione uma ou mais variáveis de controle.
- Clicar em **Estatísticas** para os testes e medidas de associação para tabelas ou subtabelas de dois fatores.
- Clicar em **Células** para obter valores observados e esperados, porcentagens e resíduos.
- Clicar em **Formatar** para controlar a ordem das categorias.

## Camadas de Crosstabs

Se você selecionar uma ou mais variáveis de camada, uma tabulação cruzada separada será produzida para cada categoria de cada variável de camada (variável de controle). Por exemplo, se você tiver uma variável de linha, uma variável de coluna e uma variável de camada com duas categorias, obterá uma tabela bidirecional para cada categoria da variável de camada. Para criar outra camada de variáveis de controle, clique em **Avançar**. As subtabelas são produzidas para cada combinação de categorias para cada variável de primeira camada, cada variável de segunda camada e assim por diante. Se forem solicitadas estatísticas e medidas de associação, elas se aplicam apenas a subtabelas bidirecionais.

## Gráfico de barras agrupadas de Crosstabs

**Exibir gráficos de barras agrupadas.** Um gráfico de barras agrupado ajuda a resumir seus dados para grupos de casos. Há um cluster de barras para cada valor da variável que você especificou em Linhas. A variável que define as barras em cada cluster é a variável que você especificou em Colunas. Há um conjunto de barras de cores ou padrões diferentes para cada valor dessa variável. Se você especificar mais de uma variável em Colunas ou Linhas, um gráfico de barras agrupado será produzido para cada combinação de duas variáveis.

## Crosstabs exibindo variáveis de camada em camadas da tabela

**Exibir variáveis de camada nas camadas da tabela.** É possível optar por exibir as variáveis de camada (variáveis de controle) como camadas de tabela na tabela de tabulação cruzada. Isso permite que você crie exibições que mostram as estatísticas gerais para variáveis de linha e coluna, bem como permite detalhar categorias de variáveis de camada.

Um exemplo que utiliza o arquivo de dados *demo.sav* (disponível no diretório de Amostras do diretório de instalação) é mostrado abaixo e foi obtido conforme a seguir:

1. Selecione *Categoria de renda em milhares (inccat)* como a variável de linha, *Possui PDA (ownpda)* como a variável de coluna e *Nível de Educação (ed)* como a variável de camada.
2. Selecione **Exibir variáveis de camada em camadas da tabela**.
3. Selecione **Coluna** no subdiálogo Exibição de Célula.
4. Execute o procedimento Crosstabs, clique duas vezes na tabela de tabulação cruzada e selecione **Ensino superior** na lista suspensa Nível de educação.

A visualização selecionada da tabela de tabulação cruzada mostra as estatísticas dos entrevistados que possuem diploma universitário.

## Estatísticas de Crosstabs

**Qui-quadrado.** Para tabelas com duas linhas e duas colunas, selecione **Qui-quadrado** para calcular o qui-quadrado de Pearson, o qui-quadrado de razão de verossimilhança, o teste exato de Fisher e o qui-quadrado corrigido de Yates (correção de continuidade). Para tabelas  $2 \times 2$ , o teste exato de Fisher é calculado quando uma tabela que não resulta de linhas ou colunas ausentes em uma tabela maior tem uma célula com uma frequência esperada menor que 5. O qui-quadrado corrigido de Yates é calculado

para todas as outras tabelas  $2 \times 2$ . Para tabelas com qualquer número de linhas e colunas, selecione **Qui-quadrado** para calcular o qui-quadrado de Pearson e o qui-quadrado de razão de verossimilhança. Quando ambas as variáveis da tabela são quantitativas, **Qui-quadrado** produz o teste de associação linear por linear.

**Correlações.** Para tabelas nas quais as linhas e colunas contêm valores ordenados, **Correlações** gera o coeficiente de correlação e o  $r$  de Spearman (somente dados numéricos). O  $r$  de Spearman é uma medida de associação entre as ordens de classificação. Quando as variáveis de tabela (fatores) são quantitativas, **Correlações** gera o coeficiente de correlação de Pearson,  $r$ , que é uma medida de associação linear entre as variáveis.

**Nominal.** Para dados nominais (nenhuma ordem intrínseca, como católica, protestante e judaica), é possível selecionar **Coefficiente de contingência, Fi** (coeficiente) e **V de Cramér, Lambda** (lambdas simétricos e assimétricos e tau de Goodman e Kruskal, e **Coefficiente de incerteza**.

- *Coefficiente de contingência.* Uma medida de associação com base em qui-quadrado. O valor varia entre 0 e 1, com 0 indicando nenhuma associação entre as variáveis de linha e de coluna e valores próximos de 1 indicando um alto grau de associação entre as variáveis. O valor máximo possível depende do número de linhas e de colunas em uma tabela.
- *Fi e V de Cramer.* Phi é uma medida de associação baseada em qui-quadrado que envolve dividir a estatística qui-quadrado pelo tamanho da amostra e obter a raiz quadrada do resultado. V de Cramer é uma medida de associação baseada em qui-quadrado.
- *Lambda.* Uma medida de associação que reflete a redução proporcional no erro quando valores da variável independente são usados para prever valores da variável dependente. Um valor de 1 significa que a variável independente prediz perfeitamente a variável dependente. Um valor 0 significa que a variável independente não ajuda a prever a variável dependente.
- *Coefficiente de incerteza.* Uma medida de associação que indica a redução proporcional no erro quando valores de uma variável são usados para prever valores da outra variável. Por exemplo, um valor de 0,83 indica que o conhecimento de uma variável reduz em 83% o erro ao prever valores da outra variável. O programa calcula ambas as versões simétrica e assimétrica do coeficiente de incerteza.

**Ordinal.** Para tabelas nas quais as linhas e colunas contêm valores ordenados, selecione **Gama** (ordem zero para tabelas de dois fatores e condicional para tabelas de 3 a 10 fatores), **Tau-b de Kendall** e **Tau-c de Kendall**. Para categorias de coluna de predição a partir de categorias de linha, selecione **d de Somers**.

- *Gamma.* Uma medida simétrica de associação entre duas variáveis ordinais que varia entre -1 e 1. Valores próximos a um valor absoluto de 1 indicam um relacionamento forte entre as duas variáveis. Valores próximos a 0 indicam pouco ou nenhum relacionamento. Para tabelas de duas vias, gamas de ordem zero são exibidas. Para tabelas de 3 a n vias, gamas condicionais são exibidos.
- *d de Somers.* Uma medida de associação entre duas variáveis ordinais que varia entre -1 e 1. Valores próximos a um valor absoluto de 1 indicam uma relação forte entre as duas variáveis, e valores próximos a 0 indicam pouca ou nenhuma relação entre as variáveis. O d de Somers é uma extensão assimétrica de gama que difere apenas na inclusão do número de pares não relacionados à variável independente. Uma versão simétrica dessa estatística também é calculada.
- *Tau-b de Kendall.* Uma medida não paramétrica de correlação para variáveis ordinais ou classificadas que levam em conta os vínculos.. O sinal do coeficiente indica a direção do relacionamento e seu valor absoluto indica a intensidade, com os valores absolutos maiores indicando os relacionamentos mais fortes. Os valores possíveis variam de -1 a 1, mas um valor de -1 ou +1 pode ser obtido apenas a partir de tabelas quadradas.
- *Tau-c de Kendall.* Uma medida de associação não paramétrica para variáveis ordinais que ignoram laços. O sinal do coeficiente indica a direção do relacionamento e seu valor absoluto indica a intensidade, com os valores absolutos maiores indicando os relacionamentos mais fortes. Os valores possíveis variam de -1 a 1, mas um valor de -1 ou +1 pode ser obtido apenas a partir de tabelas quadradas.

**Nominal por Intervalo.** Quando uma variável é categórica e a outra é quantitativa, selecione **Eta**. A variável categórica deve ser codificada de forma numérica.



- *Eta*. Uma medida de associação que varia de 0 a 1, com 0 indicando nenhuma associação entre as variáveis de linha e coluna e valores próximos de 1 indicando um alto grau de associação. O *Eta* é apropriado para uma variável dependente medida em uma escala de intervalo (por exemplo, rendimento) e para uma variável independente com um número limitado de categorias (por exemplo, sexo). Dois valores de *eta* são calculados: um trata a variável de linha como a variável de intervalo, e o outro trata a variável de coluna como a variável de intervalo.

*Kappa*. O *kappa* de Cohen mede a concordância entre as avaliações de dois avaliadores quando ambos classificam o mesmo objeto. Um valor de 1 indica concordância perfeita. Um valor de 0 indica que a concordância não é melhor que a chance. O *kappa* é baseado em uma tabela quadrada na qual os valores da linha e da coluna representam a mesma escala. Qualquer célula que tenha observado valores para uma variável mas não para outra é atribuída uma contagem de 0. A *Kappa* não é computada se o tipo de armazenamento de dados (sequência de caracteres ou numérico) não for o mesmo para as duas variáveis. Para a variável de sequência de caracteres, ambas as variáveis devem ter o mesmo comprimento definido.

*Risco*. Para tabelas 2 x 2, uma medida da intensidade da associação entre a presença de um fator e a ocorrência de um evento. Se o intervalo de confiança para a estatística incluir um valor de 1, não será possível supor que o fator está associado ao evento. A razão de chances pode ser utilizada como uma estimativa ou risco relativo quando a ocorrência do fator for rara.

*McNemar*. Um teste não paramétrico para duas variáveis dicotômicas relacionadas. Testa as mudanças nas respostas usando a distribuição qui-quadrado. Ele é útil para detectar mudanças nas respostas devido à intervenção experimental em designs de "antes e depois". Para tabelas quadradas maiores, o teste de simetria de McNemar-Bowker é relatado.

*Estatísticas de Cochran e Mantel-Haenszel*. As estatísticas de Cochran e de Mantel-Haenszel podem ser usadas para testar a independência entre uma variável de fator dicotômica e uma variável de resposta dicotômica, condicional em padrões de covariáveis definidos por uma ou mais variáveis de camada (controle).. Observe que, enquanto outras estatísticas são calculadas camada por camada, as estatísticas de Cochran e Mantel-Haenszel são calculadas uma vez para todas as camadas.

## Exibição da célula de tabulações cruzadas

Para ajudar a descobrir padrões nos dados que contribuem com um teste qui-quadrado significativo, o procedimento de tabulações cruzadas exibe as frequências esperadas e três tipos de resíduos (desvios) que medem a diferença entre as frequências observadas e esperadas. Cada célula da tabela pode conter qualquer combinação de contagens, porcentagens e resíduos selecionados.

**Contagens.** O número de casos realmente observados e o número de casos esperados se as variáveis de linha e de coluna forem independentes umas das outras. É possível optar por ocultar contagens que sejam menores que um número inteiro especificado. Os valores ocultos serão exibidos como  $< \mathbf{N}$ , onde  $\mathbf{N}$  é o inteiro especificado. O número inteiro especificado deve ser maior ou igual a 2, embora o valor 0 seja permitido e especifica que nenhuma contagem é ocultada.

**Comparar proporções da coluna.** Esta opção calcula as comparações entre pares de proporções da coluna e indica quais pares de colunas (para uma determinada linha) são significativamente diferentes. Diferenças significativas são indicadas na tabela de tabulação cruzada com formatação de estilo APA usando letras subscritas e são calculadas no nível de significância de 0,05. *Nota:* Se essa opção for especificada sem selecionar contagens observadas ou porcentagens de coluna, então as contagens observadas serão incluídas na tabela de tabulação cruzada, com as letras subscritas no estilo APA indicando os resultados dos testes de proporções da coluna.

- **Ajustar valores de p (método de Bonferroni).** Comparações entre pares de proporções da coluna utilizam a correção de Bonferroni, que ajusta o nível de significância observado para o fato de que diversas comparações são feitas.

**Porcentagens.** As porcentagens podem aumentar entre as linhas ou diminuir entre as colunas. As porcentagens do número total de casos representados na tabela (uma estrato) também estão disponíveis.

**Nota:** Se **Ocultar contagens pequenas** estiver selecionado no grupo Contagens, as porcentagens associadas a contagens ocultas também ficarão ocultas.

**Residuais.** Residuais brutos não padronizados fornecem a diferença entre os valores observados e esperados. Residuais padronizados e padronizados ajustados também estão disponíveis.

- *Não padronizado.* A diferença entre um valor observado e o valor esperado. O valor esperado é o número de casos que você esperaria na célula, se não houvesse nenhum relacionamento entre as duas variáveis. Um resíduo positivo indica que há mais casos na célula do que teria se as variáveis de linha e de coluna fossem independentes.
- *Padronizado.* O resíduo dividido por uma estimativa de seu desvio padrão. Resíduos padronizados, também conhecidos como resíduos de Pearson, possuem uma média de 0 e um desvio padrão de 1.
- *Ajustado padronizado.* O resíduo de uma célula (valor observado menos esperado) dividido por uma estimativa do seu erro padrão. O resíduo padronizado resultante é expresso em unidades de desvio padrão acima ou abaixo da média.

**Criar tabela de estilo APA.** Cria tabelas de saída que seguem as diretrizes de estilo APA.

**Nota:** As opções **Observado**, **Esperado**, **Linha**, **Coluna** e **Total** não estão disponíveis quando **Criar tabela de estilo APA** está selecionado.

**Ponderações de Não Número Inteiro.** As contagens de células são geralmente valores de números inteiros, já que eles representam o número de casos em cada célula. Mas se o arquivo de dados estiver atualmente ponderado por uma variável de ponderação com valores fracionários (por exemplo, 1,25), as contagens de células também poderão ser valores fracionários. É possível truncar ou arredondar antes ou depois de calcular as contagens de células ou utilizar contagens de células fracionárias para exibição de tabela e cálculos estatísticos.

- *Arredondar contagens de células.* As ponderações de caso são utilizadas no estado em que se encontram, e as ponderações acumuladas nas células são arredondadas antes de calcular quaisquer estatísticas.
- *Truncar contagens de células.* As ponderações de caso são utilizadas no estado em que se encontram, e as ponderações acumuladas nas células são truncadas antes de calcular quaisquer estatísticas.
- *Arredondar ponderações de caso.* As ponderações de caso são arredondadas antes do uso.
- *Truncar ponderações de caso.* As ponderações de caso são truncadas antes do uso.
- *Sem ajustes.* As ponderações de caso são usadas como estão e as contagens de células fracionárias são usadas. No entanto, quando Estatísticas exatas (disponíveis somente com Amostragem e teste) são solicitadas, as ponderações acumuladas nas células são truncadas ou arredondadas antes de calcular as estatísticas exatas do teste.

## Formato de tabela de Crosstabs

É possível organizar as linhas em ordem crescente ou decrescente dos valores da variável de linha.

## Resumir

---

O procedimento Resumir calcula estatísticas de subgrupo para variáveis dentro de categorias de uma ou mais variáveis de agrupamento. Todos os níveis da variável de agrupamento são tabulados de forma cruzada. É possível escolher a ordem em que as estatísticas são exibidas. As estatísticas básicas de cada variável em todas as categorias também são exibidas. Os valores de dados em cada categoria podem ser listados ou suprimidos. Com conjuntos de dados grandes, é possível optar por listar apenas os primeiros  $n$  casos.

**Exemplo.** Qual é o valor médio de vendas de produtos por região e setor do cliente? Você pode descobrir que o valor médio de vendas é ligeiramente maior na região oeste do que em outras regiões, com clientes corporativos na região oeste produzindo o maior valor médio de vendas.

**Estatísticas.** Soma, número de casos, média, mediana, mediana agrupada, erro padrão da média, mínimo, máximo, intervalo, valor da variável da primeira categoria da variável de agrupamento, valor da variável da última categoria da variável de agrupamento, desvio padrão, variância, curtose, erro padrão da curtose, assimetria, erro padrão da assimetria, porcentagem da soma total, porcentagem do total de  $N$ , porcentagem da soma em, porcentagem de  $N$  em, média geométrica e média harmônica.

Resumir considerações de dados

**Dados.** As variáveis de agrupamento são variáveis categóricas cujos valores podem ser numéricos ou sequência de caracteres. O número de categorias deve ser razoavelmente pequeno. As outras variáveis devem poder ser classificadas.

**Suposições.** Algumas das estatísticas de subgrupo opcionais, como a média e o desvio padrão, são baseadas na teoria normal e são apropriadas para variáveis quantitativas com distribuições simétricas. Estatísticas robustas, como a mediana e o intervalo, são apropriadas para variáveis quantitativas que podem ou não atender à suposição de normalidade.

Para Obter Sumarizações de Caso

1. A partir dos menus, escolha:

**Analisar > Relatórios > Sumarizações de Caso...**

2. Selecione duas ou mais variáveis.

Opcionalmente, é possível:

- Selecione uma ou mais variáveis de agrupamento para dividir seus dados em subgrupos.
- Clicar em **Opções** para alterar o título de saída, incluir uma legenda abaixo da saída ou excluir casos com valores omissos.
- Clicar em **Estatísticas** para estatísticas opcionais.
- Selecionar **Exibir casos** para listar os casos em cada subgrupo. Por padrão, o sistema lista apenas os primeiros 100 casos em seu arquivo. É possível aumentar ou diminuir o valor para **Limitar casos aos primeiros  $n$**  ou cancele a seleção desse item na lista todos os casos.

## Opções de sumarização

Resumir permite que você mude o título de sua saída ou adicione uma legenda que aparecerá abaixo da tabela de saída. É possível controlar a quebra de linha em títulos e legendas digitando \n onde desejar inserir uma quebra de linha no texto.

Também é possível optar por exibir ou suprimir subtítulos para totais e incluir ou excluir casos com valores omissos para qualquer uma das variáveis usadas em qualquer uma das análises. Muitas vezes, é desejável denotar casos ausentes na saída com um ponto ou um asterisco. Insira um caractere, frase ou código que você gostaria que aparecesse quando um valor estiver ausente; caso contrário, nenhum tratamento especial será aplicado aos casos ausentes na saída.

## Estatísticas de sumarização

É possível escolher uma ou mais das seguintes estatísticas de subgrupo para as variáveis dentro de cada categoria de cada variável de agrupamento: soma, número de casos, média, mediana, mediana agrupada, erro padrão da média, mínimo, máximo, intervalo, valor da variável da primeira categoria da variável de agrupamento, valor da variável da última categoria da variável de agrupamento, desvio padrão, variância, curtose, erro padrão da curtose, assimetria, erro padrão da assimetria, porcentagem da soma total, porcentagem do total de  $N$ , porcentagem da soma em, porcentagem de  $N$  em, média geométrica e média harmônica. A ordem na qual as estatísticas aparecem na lista Estatísticas da célula é a ordem na qual elas serão exibidas na saída. Estatísticas resumidas também são exibidas para cada variável em todas as categorias.

*First.* Exibe o primeiro valor de dados encontrado no arquivo de dados.

*Média Geométrica.* A  $n$ -ésima raiz do produto dos valores de dados, em que  $n$  representa o número de casos.

*Mediana Agrupada.* Mediana que é calculada para dados codificados em grupos. Por exemplo, com dados de idade, se cada valor nos anos 30 for codificado como 35, cada valor nos anos 40 for codificado como 45, e assim por diante, a mediana agrupada será a média calculada a partir dos dados codificados.

*média harmônica.* Usado para estimar um tamanho médio do grupo quando os tamanhos de amostra nos grupos não são iguais.. A média harmônica é o número total de amostras dividido pela soma dos recíprocos dos tamanhos da amostra.

*Curtose.* Uma medida da extensão até a qual há valores discrepantes Para a uma distribuição normal, o valor da estatística da curtose é zero. A curtose positiva indica que os dados exibem mais valores discrepantes que uma distribuição normal. A curtose negativa indica que os dados exibem menos valores discrepantes extremos do que uma distribuição normal. A definição de curtose que é usada, em que o valor é de 0 para uma distribuição normal, é, às vezes, referida como excesso de curtose. Alguns softwares podem reportar curtose tal que o valor é 3 para uma distribuição normal.

*Last.* Exibe o último valor de dados encontrado no arquivo de dados.

*Máximo.* O maior valor de uma variável numérica.

*Média.* Uma medida de tendência central A média aritmética, a soma dividida pelo número de casos.

*Median.* O valor acima e abaixo do qual metade dos casos cai, o 50th percentil. Se houver um número par de casos, a mediana é a média dos dois casos intermediários quando estão classificados em ordem crescente ou decrescente. A mediana é uma medida da tendência central não sensível a valores excessivos (ao contrário da média, que pode ser afetada por alguns valores extremamente altos ou baixos).

*Mínimo.* O menor valor de uma variável numérica.

*N.* O número de casos (observações ou registros).

*Percentual do N Total.* Porcentagem do número total de casos em cada categoria.

*Percentual da Soma total.* Porcentagem da soma total em cada categoria.

*Intervalo.* A diferença entre os valores maior e menor de uma variável numérica, o máximo menos o mínimo.

*Assimetria.* Uma medida da assimetria de uma distribuição A distribuição normal é simétrica e tem um valor de assimetria de 0. Uma distribuição com assimetria positiva significativa tem um longo rodapé direito. Uma distribuição com assimetria negativa significativa tem um longo rodapé esquerdo. Como uma orientação, um valor de assimetria mais de duas vezes seu erro padrão é obtido para indicar uma partida de simetria.

*desvio padrão .* Medida de dispersão em torno da média. Em uma distribuição normal, 68% dos casos estão dentro de um desvio padrão da média e 95% dos casos estão dentro de dois desvios padrão. Por exemplo, se a idade média for 45, com um desvio padrão de 10, 95% dos casos estariam entre 25 e 65 em uma distribuição normal.

*Erro Padrão de Curtose.* A razão de curtose para seu erro padrão pode ser usada como um teste de normalidade (ou seja, é possível rejeitar a normalidade se a razão for menor que -2 ou maior que + 2)... Um valor positivo grande para curtose indica que os rodapés da distribuição são maiores do que aqueles de uma distribuição normal, e um valor negativo para curtose indica rodapés mais curtos (como aqueles de uma distribuição uniforme em forma de caixa).

*Erro padrão de média.* Uma medida de quanto o valor da média pode variar de amostra para amostra obtida da mesma distribuição. Ela pode ser utilizada para comparar aproximadamente a média observada com um valor hipotético (ou seja, é possível concluir que os dois valores serão diferentes se a razão da diferença com o erro padrão for inferior a -2 ou superior a +2).

*Erro Padrão de Assimetria.* A razão de assimetria para seu erro padrão pode ser usada como um teste de normalidade (ou seja, é possível rejeitar a normalidade se a razão for menor que -2 ou maior que + 2)... Um valor positivo grande para assimetria indica um rodapé direito longo, e um valor negativo extremo indica um rodapé esquerdo longo.

*Sum.* A soma ou o total dos valores, em todos os casos com valores não omissos.

*Variância.* Uma medida de dispersão ao redor da média, igual à soma dos desvios quadrados da média dividido por um menor que o número de casos. A variância é medida em unidades que são o quadrado daquelas da própria variável.

## Médias

O procedimento Médias calcula médias de subgrupo e estatísticas univariadas relacionadas para variáveis dependentes dentro das categorias de uma ou mais variáveis independentes. Opcionalmente, é possível obter uma análise de variância para um fator, eta e testes para linearidade.

**Exemplo.** Avalie a quantia média de gordura absorvida por três tipos diferentes de óleo de cozinha, e execute uma análise de variância para um fator para ver se as médias diferem.

**Estatísticas.** Soma, número de casos, média, mediana, mediana agrupada, erro padrão da média, mínimo, máximo, intervalo, valor da variável da primeira categoria da variável de agrupamento, valor da variável da última categoria da variável de agrupamento, desvio padrão, variância, curtose, erro padrão da curtose, assimetria, erro padrão da assimetria, porcentagem da soma total, porcentagem do total de  $N$ , porcentagem da soma em, porcentagem de  $N$  em, média geométrica e média harmônica. As opções incluem análise de variância, eta, eta quadrado, e testes para linearidade  $R$  e  $R^2$ .

Considerações de dados de médias

**Dados.** As variáveis dependentes são quantitativas, e as variáveis independentes são categóricas. Os valores de variáveis categóricas podem ser numéricos ou sequência de caracteres.

**Suposições.** Algumas das estatísticas de subgrupo opcionais, como a média e o desvio padrão, baseiam-se em teoria normal e são apropriadas para variáveis quantitativas com distribuições simétricas. Estatísticas robustas, como a mediana, são apropriadas para variáveis quantitativas que podem ou não satisfazer a suposição de normalidade. A análise de variância é robusta para partidas da normalidade, mas os dados em cada célula devem ser simétricos. A análise de variância também supõe que os grupos vêm de populações com variâncias iguais. Para testar essa suposição, use o teste de homogeneidade de variância de Levene, disponível no procedimento One-Way ANOVA.

Para obter médias de subgrupo

1. A partir dos menus, escolha:

**Analisar > Comparar Médias > Médias...**

2. Selecione uma ou mais variáveis dependentes.
3. Use um dos seguintes métodos para selecionar variáveis independentes categóricas:
  - Selecione uma ou mais variáveis independentes. Resultados separados são exibidos para cada variável independente.
  - Selecione uma ou mais estratos para variáveis independentes. Cada camada subdivide ainda mais a amostra. Se você tiver uma variável independente na Camada 1 e uma variável independente na Camada 2, os resultados serão exibidos em uma tabela cruzada, ao invés de tabelas separadas para cada variável independente.
4. Opcionalmente, clique em **Opções** para estatísticas opcionais, uma análise de tabela de variância, eta, eta quadrado,  $R$  e  $R^2$ .

## Opções de Médias

É possível escolher uma ou mais das seguintes estatísticas de subgrupo para as variáveis dentro de cada categoria de cada variável de agrupamento: soma, número de casos, média, mediana, mediana agrupada, erro padrão da média, mínimo, máximo, intervalo, valor da variável da primeira categoria da variável de agrupamento, valor da variável da última categoria da variável de agrupamento, desvio padrão, variância, curtose, erro padrão da curtose, assimetria, erro padrão da assimetria, porcentagem da soma total, porcentagem do total de  $N$ , porcentagem da soma em, porcentagem de  $N$  em, média geométrica e média harmônica. É possível mudar a ordem em que as estatísticas do subgrupo aparecem. A ordem na qual as estatísticas aparecem na lista Estatísticas da Célula é a ordem na qual elas são exibidas na saída. Estatísticas resumidas também são exibidas para cada variável em todas as categorias.

*First.* Exibe o primeiro valor de dados encontrado no arquivo de dados.

*Média Geométrica.* A  $n$ -ésima raiz do produto dos valores de dados, em que  $n$  representa o número de casos.

*Mediana Agrupada.* Mediana que é calculada para dados codificados em grupos. Por exemplo, com dados de idade, se cada valor nos anos 30 for codificado como 35, cada valor nos anos 40 for codificado como 45, e assim por diante, a mediana agrupada será a média calculada a partir dos dados codificados.

*média harmônica.* Usado para estimar um tamanho médio do grupo quando os tamanhos de amostra nos grupos não são iguais.. A média harmônica é o número total de amostras dividido pela soma dos recíprocos dos tamanhos da amostra.

*Curtose.* Uma medida da extensão até a qual há valores discrepantes Para a uma distribuição normal, o valor da estatística da curtose é zero. A curtose positiva indica que os dados exibem mais valores discrepantes que uma distribuição normal. A curtose negativa indica que os dados exibem menos valores discrepantes extremos do que uma distribuição normal. A definição de curtose que é usada, em que o valor é de 0 para uma distribuição normal, é, às vezes, referida como excesso de curtose. Alguns softwares podem reportar curtose tal que o valor é 3 para uma distribuição normal.

*Last.* Exibe o último valor de dados encontrado no arquivo de dados.

*Máximo.* O maior valor de uma variável numérica.

*Média.* Uma medida de tendência central A média aritmética, a soma dividida pelo número de casos.

*Median.* O valor acima e abaixo do qual metade dos casos cai, o 50th percentil. Se houver um número par de casos, a mediana é a média dos dois casos intermediários quando estão classificados em ordem crescente ou decrescente. A mediana é uma medida da tendência central não sensível a valores excessivos (ao contrário da média, que pode ser afetada por alguns valores extremamente altos ou baixos).

*Mínimo.* O menor valor de uma variável numérica.

*N.* O número de casos (observações ou registros).

*Percentual de N total.* Porcentagem do número total de casos em cada categoria.

*Percentual da soma total.* Porcentagem da soma total em cada categoria.

*Intervalo.* A diferença entre os valores maior e menor de uma variável numérica, o máximo menos o mínimo.

*Assimetria.* Uma medida da assimetria de uma distribuição A distribuição normal é simétrica e tem um valor de assimetria de 0. Uma distribuição com assimetria positiva significativa tem um longo rodapé direito. Uma distribuição com assimetria negativa significativa tem um longo rodapé esquerdo. Como uma orientação, um valor de assimetria mais de duas vezes seu erro padrão é obtido para indicar uma partida de simetria.

*desvio padrão .* Medida de dispersão em torno da média. Em uma distribuição normal, 68% dos casos estão dentro de um desvio padrão da média e 95% dos casos estão dentro de dois desvios padrão. Por exemplo, se a idade média for 45, com um desvio padrão de 10, 95% dos casos estariam entre 25 e 65 em uma distribuição normal.

*Erro Padrão de Curtose.* A razão de curtose para seu erro padrão pode ser usada como um teste de normalidade (ou seja, é possível rejeitar a normalidade se a razão for menor que -2 ou maior que + 2)... Um valor positivo grande para curtose indica que os rodapés da distribuição são maiores do que aqueles de uma distribuição normal, e um valor negativo para curtose indica rodapés mais curtos (como aqueles de uma distribuição uniforme em forma de caixa).

*Erro padrão de média.* Uma medida de quanto o valor da média pode variar de amostra para amostra obtida da mesma distribuição. Ela pode ser utilizada para comparar aproximadamente a média observada com um valor hipotético (ou seja, é possível concluir que os dois valores serão diferentes se a razão da diferença com o erro padrão for inferior a -2 ou superior a +2).

*Erro Padrão de Assimetria.* A razão de assimetria para seu erro padrão pode ser usada como um teste de normalidade (ou seja, é possível rejeitar a normalidade se a razão for menor que -2 ou maior que + 2)... Um valor positivo grande para assimetria indica um rodapé direito longo, e um valor negativo extremo indica um rodapé esquerdo longo.

*Sum.* A soma ou o total dos valores, em todos os casos com valores não omissos.

*Variância.* Uma medida de dispersão ao redor da média, igual à soma dos desvios quadrados da média dividido por um menor que o número de casos. A variância é medida em unidades que são o quadrado daquelas da própria variável.

Estatísticas para a Primeira Camada

*Tabela anova e eta.* Exibe uma tabela de análise de variância unidirecional e calcula o eta e o eta quadrado (medidas de associação) para cada variável independente na primeira camada.

*Teste para linearidade.* Calcula a soma de quadrados, graus de liberdade e quadrados médios associados a componentes lineares e não lineares, bem como a razão F, R e R ao quadrado A linearidade não será calculada se a variável independente for uma sequência de caracteres curta.

## Cubos OLAP

---

O procedimento Cubos OLAP (Processamento Analítico On-line) calcula totais, médias e outras estatísticas univariadas para variáveis de resumo contínuas dentro de categorias de uma ou mais variáveis de agrupamento categórico. Uma camada separada na tabela é criada para cada categoria de cada variável de agrupamento.

**Exemplo.** Vendas totais e médias para diferentes regiões e linhas de produtos dentro das regiões.

**Estatísticas.** Soma, número de casos, média, mediana, mediana agrupada, erro padrão da média, mínimo, máximo, intervalo, valor da variável da primeira categoria da variável de agrupamento, valor da variável da última categoria da variável de agrupamento, desvio padrão, variância, curtose, erro padrão de curtose, assimetria, erro padrão de assimetria, porcentagem de casos totais, porcentagem de soma total, porcentagem de casos totais dentro de variáveis de agrupamento, porcentagem de soma total dentro de variáveis de agrupamento, média geométrica e média harmônica.

Considerações sobre dados de cubos OLAP

**Dados.** As variáveis de resumo são quantitativas (variáveis contínuas medidas em uma escala de intervalo ou razão), e as variáveis de agrupamento são categóricas. Os valores de variáveis categóricas podem ser numéricos ou sequência de caracteres.

**Suposições.** Algumas das estatísticas de subgrupo opcionais, como a média e o desvio padrão, são baseadas na teoria normal e são apropriadas para variáveis quantitativas com distribuições simétricas. Estatísticas robustas, como a mediana e o intervalo, são apropriadas para variáveis quantitativas que podem ou não atender à suposição de normalidade.

Para obter cubos OLAP

1. No menu, escolha:

**Analisar > Relatórios > Cubos OLAP...**

2. Selecione uma ou mais variáveis de sumarização contínuas.

3. Selecione uma ou mais variáveis de agrupamento categóricas.

Opcionalmente:

- Selecione estatísticas básicas diferentes (clique em **Estatísticas**). Deve-se selecionar uma ou mais variáveis de agrupamento antes de poder selecionar estatísticas de resumo.
- Calcule diferenças entre os pares de variáveis e pares de grupos que estiverem definidos por uma variável de agrupamento (clique em **Diferenças**).
- Crie títulos de tabela customizados (clique em **Título**).
- Oculte contagens que forem menores que um número inteiro especificado. Os valores ocultos serão exibidos como **<N**, em que **N** é o número inteiro especificado. O número inteiro especificado deve ser maior ou igual a 2.

## Estatísticas de Cubos OLAP

É possível escolher uma ou mais das seguintes estatísticas de subgrupo para as variáveis de resumo dentro de cada categoria de cada variável de agrupamento: soma, número de casos, média, mediana, mediana agrupada, erro padrão da média, mínimo, máximo, intervalo, valor da variável da primeira categoria da variável de agrupamento, valor da variável da última categoria da variável de agrupamento, desvio padrão, variância, curtose, erro padrão de curtose, assimetria, erro padrão de assimetria, porcentagem do total de casos, porcentagem da soma total, porcentagem de casos totais dentro de variáveis de agrupamento, porcentagem da soma total dentro de variáveis de agrupamento, média geométrica e média harmônica.

É possível mudar a ordem em que as estatísticas do subgrupo aparecem. A ordem na qual as estatísticas aparecem na lista Estatísticas da célula é a ordem na qual elas são exibidas na saída. Estatísticas resumidas também são exibidas para cada variável em todas as categorias.

*First.* Exibe o primeiro valor de dados encontrado no arquivo de dados.

*Média Geométrica.* A  $n$ -ésima raiz do produto dos valores de dados, em que  $n$  representa o número de casos.

*Mediana Agrupada.* Mediana que é calculada para dados codificados em grupos. Por exemplo, com dados de idade, se cada valor nos anos 30 for codificado como 35, cada valor nos anos 40 for codificado como 45, e assim por diante, a mediana agrupada será a média calculada a partir dos dados codificados.

*média harmônica.* Usado para estimar um tamanho médio do grupo quando os tamanhos de amostra nos grupos não são iguais.. A média harmônica é o número total de amostras dividido pela soma dos recíprocos dos tamanhos da amostra.

*Curtose.* Uma medida da extensão até a qual há valores discrepantes Para a uma distribuição normal, o valor da estatística da curtose é zero. A curtose positiva indica que os dados exibem mais valores discrepantes que uma distribuição normal. A curtose negativa indica que os dados exibem menos valores discrepantes extremos do que uma distribuição normal. A definição de curtose que é usada, em que o valor é de 0 para uma distribuição normal, é, às vezes, referida como excesso de curtose. Alguns softwares podem reportar curtose tal que o valor é 3 para uma distribuição normal.

*Last.* Exibe o último valor de dados encontrado no arquivo de dados.

*Máximo.* O maior valor de uma variável numérica.

*Média.* Uma medida de tendência central A média aritmética, a soma dividida pelo número de casos.

*Median.* O valor acima e abaixo do qual metade dos casos cai, o 50th percentil. Se houver um número par de casos, a mediana é a média dos dois casos intermediários quando estão classificados em ordem crescente ou decrescente. A mediana é uma medida da tendência central não sensível a valores excessivos (ao contrário da média, que pode ser afetada por alguns valores extremamente altos ou baixos).

*Mínimo.* O menor valor de uma variável numérica.

*N.* O número de casos (observações ou registros).

*Percentual de N em.* Porcentagem do número de casos para a variável de agrupamento especificada em categorias de outras variáveis de agrupamento. Se você tiver apenas uma variável de agrupamento, esse valor será idêntico à porcentagem do número total de casos.

*Percentual de soma em.* Porcentagem da soma para a variável de agrupamento especificada nas categorias de outras variáveis de agrupamento. Se você tiver apenas uma variável de agrupamento, esse valor será idêntico à porcentagem da soma total.

*Percentual do N Total.* Porcentagem do número total de casos em cada categoria.

*Percentual da Soma total.* Porcentagem da soma total em cada categoria.

*Intervalo.* A diferença entre os valores maior e menor de uma variável numérica, o máximo menos o mínimo.



*Assimetria.* Uma medida da assimetria de uma distribuição A distribuição normal é simétrica e tem um valor de assimetria de 0. Uma distribuição com assimetria positiva significativa tem um longo rodapé direito. Uma distribuição com assimetria negativa significativa tem um longo rodapé esquerdo. Como uma orientação, um valor de assimetria mais de duas vezes seu erro padrão é obtido para indicar uma partida de simetria.

*desvio padrão .* Medida de dispersão em torno da média. Em uma distribuição normal, 68% dos casos estão dentro de um desvio padrão da média e 95% dos casos estão dentro de dois desvios padrão. Por exemplo, se a idade média for 45, com um desvio padrão de 10, 95% dos casos estariam entre 25 e 65 em uma distribuição normal.

*Erro Padrão de Curtose.* A razão de curtose para seu erro padrão pode ser usada como um teste de normalidade (ou seja, é possível rejeitar a normalidade se a razão for menor que -2 ou maior que + 2)... Um valor positivo grande para curtose indica que os rodapés da distribuição são maiores do que aqueles de uma distribuição normal, e um valor negativo para curtose indica rodapés mais curtos (como aqueles de uma distribuição uniforme em forma de caixa).

*Erro padrão de média.* Uma medida de quanto o valor da média pode variar de amostra para amostra obtida da mesma distribuição. Ela pode ser utilizada para comparar aproximadamente a média observada com um valor hipotético (ou seja, é possível concluir que os dois valores serão diferentes se a razão da diferença com o erro padrão for inferior a -2 ou superior a +2).

*Erro Padrão de Assimetria.* A razão de assimetria para seu erro padrão pode ser usada como um teste de normalidade (ou seja, é possível rejeitar a normalidade se a razão for menor que -2 ou maior que + 2)... Um valor positivo grande para assimetria indica um rodapé direito longo, e um valor negativo extremo indica um rodapé esquerdo longo.

*Sum.* A soma ou o total dos valores, em todos os casos com valores não omissos.

*Variância.* Uma medida de dispersão ao redor da média, igual à soma dos desvios quadrados da média dividido por um menor que o número de casos. A variância é medida em unidades que são o quadrado daquelas da própria variável.

## Diferenças de Cubos OLAP

Essa caixa de diálogo permite calcular diferenças percentuais e aritméticas entre variáveis de resumo ou entre grupos definidos por uma variável de agrupamento. As diferenças são calculadas para todas as medidas selecionadas na caixa de diálogo Estatísticas de Cubos OLAP.

**Diferenças entre Variáveis.** Calcula diferenças entre pares de variáveis. Os valores de estatísticas básicas para a segunda variável (a variável Menos) em cada par são subtraídos dos valores de estatísticas de resumo para a primeira variável do par. Para diferenças percentuais, o valor da variável de resumo para a variável Menos é usado como denominador. Deve-se selecionar pelo menos duas variáveis de resumo na caixa de diálogo principal antes de poder especificar as diferenças entre as variáveis.

**Diferenças entre Grupos de Casos.** Calcula diferenças entre pares de grupos definidos por uma variável de agrupamento. Os valores de estatísticas básicas para a segunda categoria em cada par (a categoria Menos) são subtraídos dos valores de estatísticas básicas para a primeira categoria do par. As diferenças percentuais usam o valor da estatística básica para a categoria Menos como denominador. Deve-se selecionar uma ou mais variáveis de agrupamento na caixa de diálogo principal antes de especificar as diferenças entre os grupos.

## Título de Cubos OLAP

É possível alterar o título de sua saída ou incluir uma legenda que aparecerá abaixo da tabela de saída. Também é possível controlar a quebra de linha em títulos e legendas ao digitar \n onde desejar inserir uma quebra de linha no texto.

## Introdução de proporções

As proporções calculam testes e intervalos de confiança para proporções ou diferenças binomiais de proporções. As estatísticas estão disponíveis para as proporções de uma amostra (testadas com relação a um valor especificado), amostras emparelhadas (variáveis diferentes) ou amostras independentes (grupos de casos diferentes). Várias opções para tipos de estatísticas de teste e intervalos de confiança são oferecidas. Outros procedimentos que fornecem algumas das mesmas funcionalidades incluem **CROSSTABS**, **NPAR TESTS** e **NPTESTS**.

### Proporções de uma amostra

Testes de uma amostra e intervalos de confiança. A saída inclui a proporção observada, a estimativa da diferença entre a proporção populacional e a proporção populacional hipotética, erros padrão assintóticos sob hipóteses nulas e alternativas, estatísticas de teste especificadas com probabilidades bilaterais e intervalos de confiança especificados para proporções.

### Proporções de amostras pareadas

Testes de amostras emparelhadas e intervalos de confiança para diferenças entre as proporções. A saída inclui proporções observadas, estimativas de diferenças de proporções populacionais, erros de padrão assintóticos de diferenças populacionais sob hipóteses nulas e alternativas, estatísticas de teste especificadas com probabilidades bilaterais e intervalos de confiança especificados para diferenças em proporções.

### Proporções de amostras independentes

Testes de amostras independentes e intervalos de confiança. A saída inclui proporções observadas, estimativas de diferenças de proporções populacionais, erros de padrão assintóticos de diferenças populacionais sob hipóteses nulas e alternativas, estatísticas de teste especificadas com probabilidades bilaterais e intervalos de confiança especificados para diferenças em proporções.

## Proporções de uma amostra

O procedimento One-Sample Proportions fornece testes e intervalos de confiança para proporções binomiais individuais. Os dados são assumidos como sendo de uma amostra aleatória simples, e cada teste de hipótese ou intervalo de confiança é um teste separado ou intervalo individual, baseado em uma proporção binomial. A saída inclui a proporção observada, a estimativa da diferença entre a proporção da população e a proporção da população hipotética, erros padrão assintóticos sob hipóteses nulas e alternativas, estatísticas de teste especificadas com probabilidades bilaterais e intervalos de confiança especificados para proporções.

### Exemplo

#### Estatísticas

Agresti-Coull, Anscombe, Clopper-Pearson (Exato), Jeffreys, Logit, Wald, Wald (continuidade corrigida), Wilson Score, Wilson Score (continuidade corrigida), Exato Binomial, Mid-p Adjusted Binomial, Score, Score (continuidade corrigida).

### Considerações de dados

#### Dados

O procedimento exibe estatísticas de teste solicitadas e probabilidades bilaterais, intervalos de confiança para diferenças de proporções, bem como proporções, erros padrão e contagens para cada grupo ou variável. O procedimento é limitado ao máximo de um valor de teste.

#### Suposições

### Obtendo testes de proporções de uma amostra

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de uma amostra...**

2. Selecione uma ou mais variáveis de teste quantitativas.

3. Opcionalmente, é possível:

- Selecione as configurações de critérios de sucesso na seção **Definir sucesso**:

**Último valor**

O último ou o maior valor entre os valores distintos classificados nos dados é usado. Isso se aplica a variáveis numéricas ou de sequência. Essa é a configuração padrão.

**Primeiro valor**

O primeiro ou o menor valor entre os valores distintos classificados nos dados é usado. Isso se aplica a variáveis numéricas ou de sequência.

**Valor(es)**

Um ou mais valores específicos entre parênteses Vários valores devem ser separados por espaços. Isso se aplica a variáveis numéricas ou de sequência. Os valores das variáveis de string devem ser colocados entre aspas simples.

**Ponto médio**

Os valores no ou acima do meio do intervalo de valores observados nos dados. Isso se aplica apenas a dados numéricos.

**Ponto de Corte**

Valores no ou acima de um valor especificado. Isso se aplica apenas a dados numéricos.

- Clique em **Intervalos de confiança...** para especificar quais tipos de intervalos de confiança são exibidos ou para suprimir todos os intervalos de confiança.
- Clique em **Testes...** para especificar quais tipos de estatísticas de teste são exibidas ou para suprimir todos os testes.
- Clique em **Valores omissos...** para controlar o tratamento de dados omissos.
- Clique em **Autoinicialização...** para derivar estimativas robustas de erros padrão e intervalos de confiança para estimativas como a média, mediana, proporção, razão de chances, coeficiente de correlação ou coeficiente de regressão.

4. Clique em **OK**.

## Proporções de uma amostra: intervalos de confiança

A caixa de diálogo Intervalos de confiança fornece opções para especificar o nível de cobertura e para selecionar quais tipos de intervalos de confiança são exibidos.

**Nível de Cobertura**

Especifica a porcentagem do intervalo de confiança. Um valor numérico no intervalo (0,100) deve ser especificado. 95 é a configuração padrão.

**Tipo(s) de intervalo**

Fornece opções para especificar quais tipos de intervalos de confiança são exibidos. As opções disponíveis incluem:

- Agresti-Coull
- Anscombe
- Clopper-Pearson (Exato)
- Jeffreys
- Logit
- Wald
- Wald (Continuidade corrigida)
- Pontuação de Wilson
- Pontuação de Wilson (continuidade corrigida)

## Especificando intervalos de confiança para proporções de uma amostra

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de uma amostra...**

2. Clique em **Intervalos de confiança** para especificar quais tipos de intervalos de confiança são exibidos ou para suprimir todos os intervalos de confiança.

## Proporções de uma amostra: testes

O diálogo Testes fornece opções para especificar quais tipos de estatísticas de teste são exibidos.

### Todos

Todas as estatísticas de teste são exibidas na saída.

### Nenhum

Nenhuma estatística de teste é exibida na saída.

### Binômio exato

Exibe probabilidades binomiais exatas.

### Binômio ajustado de meio-p

Exibe as probabilidades binomiais ajustadas para mid-p. Esta é uma configuração padrão.

### Pontuação

Exibe a estatística de teste do Score Z. Esta é uma configuração padrão.

### Pontuação (Continuidade corrigida)

Exibe a estatística de teste Z de pontuação corrigida pela continuidade.

### Wald

Exibe a estatística de teste Wald Z.

### Wald (Continuidade corrigida)

Exibe a estatística de teste Wald Z corrigida pela continuidade.

### Valor de Teste

Especifica um valor de teste entre 0 e 1. O valor padrão é 0.5.

## Obtendo testes de proporções de uma amostra

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de uma amostra...**

2. No diálogo Proporções de uma amostra, clique em **Testes**.

3. Selecione um ou mais dos testes disponíveis.

## Proporções de uma amostra: valores omissos

O diálogo Valores omissos fornece opções para lidar com valores omissos.

### Escopo de dados ausentes

#### Excluir análise de casos por análise

Indica a inclusão de todos os casos com dados suficientes sobre as variáveis usadas em cada análise particular. Essa é a configuração padrão.

#### Excluir casos de listwise

Indica a inclusão de todos os casos com dados suficientes sobre todas as variáveis usadas em todas as análises.

### Valores ausentes do usuário

**Excluir** trata os valores omissos do usuário como omissos. **Incluir** ignora designações de valor omissos do usuário e trata os valores omissos do usuário como válidos.

## Definindo configurações de valor omissos para Proporções de uma amostra

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de uma amostra...**

2. No diálogo Proporções de uma amostra, clique em **Valores omissos**.

3. Selecione as configurações de valores omissos desejadas.

## Proporções de amostras pareadas

O procedimento de proporções de amostras pareadas fornece testes e intervalos de confiança para a diferença em duas proporções binomiais relacionadas ou emparelhadas. Os dados são assumidos como sendo de uma amostra aleatória simples, e cada teste de hipótese ou intervalo de confiança é um teste separado ou intervalo individual. A saída inclui proporções observadas, estimativas de diferenças em proporções populacionais, erros padrão assintóticos de diferenças populacionais sob hipóteses nulas e alternativas, estatísticas de teste especificadas com probabilidades bilaterais e intervalos de confiança especificados para diferenças em proporções.

### Exemplo

#### Estatísticas

Agresti-Min, Bonett-Price, Newcombe, Wald, Wald (continuidade corrigida), Binomial Exata, Mid-p Binomial Ajustada, McNemar, McNemar (continuidade corrigida).

## Considerações de dados

### Dados

- Uma lista de variáveis contendo pelo menos duas variáveis é necessária.
- Se uma única lista de variáveis for especificada, cada membro da lista será emparelhado com todos os outros membros da lista.

### Suposições

- Se duas listas de variáveis forem separadas por **COM** sem a palavra-chave (**PAIRED**), cada membro da primeira lista será emparelhado com cada membro da segunda lista.
- Se duas listas de variáveis forem separadas por **WITH** e a segunda lista for seguida por (**PAIRED**), os membros das duas listas em ordem serão emparelhados: o primeiro membro da primeira lista será emparelhado com o primeiro membro da segunda lista, o segundo membros de cada lista são emparelhados, etc. Variáveis sem correspondência são ignoradas e uma mensagem de aviso é emitida.

## Obtendo testes de proporções de amostras emparelhadas

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de amostras emparelhadas...**

2. Selecione uma ou mais variáveis de teste quantitativas.

3. Opcionalmente, é possível:

- Selecione as configurações de critérios de sucesso na seção **Definir sucesso**:

#### Último valor

O último ou o maior valor entre os valores distintos classificados nos dados é usado. Isso se aplica a variáveis numéricas ou de sequência. Essa é a configuração padrão.

#### Primeiro valor

O primeiro ou o menor valor entre os valores distintos classificados nos dados é usado. Isso se aplica a variáveis numéricas ou de sequência.

**Valor(es)**

Um ou mais valores específicos entre parênteses Vários valores devem ser separados por espaços. Isso se aplica a variáveis numéricas ou de sequência. Os valores das variáveis de string devem ser colocados entre aspas simples.

**Ponto médio**

Os valores no ou acima do meio do intervalo de valores observados nos dados. Isso se aplica apenas a dados numéricos.

**Ponto de Corte**

Valores no ou acima de um valor especificado. Isso se aplica apenas a dados numéricos.

- Clique em **Intervalos de confiança...** para especificar quais tipos de intervalos de confiança são exibidos ou para suprimir todos os intervalos de confiança.
- Clique em **Testes...** para especificar quais tipos de estatísticas de teste são exibidas ou para suprimir todos os testes.
- Clique em **Valores omissos...** para controlar o tratamento de dados omissos.
- Clique em **Autoinicialização...** para derivar estimativas robustas de erros padrão e intervalos de confiança para estimativas como a média, mediana, proporção, razão de chances, coeficiente de correlação ou coeficiente de regressão.

4. Clique em **OK**.

## Proporções de amostras emparelhadas: intervalos de confiança

A caixa de diálogo Intervalos de confiança fornece opções para especificar o nível de cobertura e para selecionar quais tipos de intervalos de confiança são exibidos.

**Nível de Cobertura**

Especifica a porcentagem do intervalo de confiança. Um valor numérico no intervalo (0,100) deve ser especificado. 95 é a configuração padrão.

**Tipo(s) de intervalo**

Fornecer opções para especificar quais tipos de intervalos de confiança são exibidos. As opções disponíveis incluem:

- Agresti-Min
- Bonett-Price
- Newcombe
- Wald
- Wald (Continuidade corrigida)

## Especificando intervalos de confiança para proporções de amostras emparelhadas

1. No menu, escolha:

**Analisar > Comparar médias > Proporções de amostras emparelhadas...**

2. Clique em **Intervalos de confiança** para especificar quais tipos de intervalos de confiança são exibidos ou para suprimir todos os intervalos de confiança.

## Proporções de amostras emparelhadas: testes

O diálogo Testes fornece opções para especificar quais tipos de estatísticas de teste são exibidos.

**Todos**

Todas as estatísticas de teste são exibidas na saída.

**Nenhum**

Nenhuma estatística de teste é exibida na saída.

**Binômio exato**

Exibe probabilidades binomiais exatas.

### **Binômio ajustado de meio-p**

Exibe as probabilidades binomiais ajustadas para mid-p. Esta é uma configuração padrão.

### **McNemar**

Exibe a estatística de teste do McNemar Z. Esta é uma configuração padrão.

### **McNemar (Continuidade corrigida)**

Exibe a estatística de teste McNemar Z corrigida pela continuidade.

### **Wald**

Exibe a estatística de teste Wald Z.

### **Wald (Continuidade corrigida)**

Exibe a estatística de teste Wald Z corrigida pela continuidade.

## **Obtendo testes de proporções de amostras emparelhadas**

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de amostras emparelhadas...**

2. No diálogo Proporções de amostras emparelhadas, clique em **Testes**.

3. Selecione um ou mais dos testes disponíveis.

## **Proporções de amostras emparelhadas: valores omissos**

O diálogo Valores omissos fornece opções para lidar com valores omissos.

### **Escopo de dados ausentes**

#### **Excluir análise de casos por análise**

Indica a inclusão de todos os casos com dados suficientes sobre as variáveis usadas em cada análise particular. Essa é a configuração padrão.

#### **Excluir casos de listwise**

Indica a inclusão de todos os casos com dados suficientes sobre todas as variáveis usadas em todas as análises.

### **Valores ausentes do usuário**

**Excluir** trata os valores omissos do usuário como omissos. **Incluir** ignora designações de valor omissos do usuário e trata os valores omissos do usuário como válidos.

## **Definindo configurações de valor omissos para Proporções de amostras emparelhadas**

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de amostras emparelhadas...**

2. No diálogo Proporções de amostras emparelhadas, clique em **Valores omissos**.

3. Selecione as configurações de valores omissos desejadas.

## **Proporções de amostras independentes**

O procedimento de Proporções de Amostras Independentes fornece testes e intervalos de confiança para a diferença em duas proporções binomiais independentes. Os dados são assumidos como sendo de uma amostra aleatória simples, e cada teste de hipótese ou intervalo de confiança é um teste separado ou intervalo individual. A saída inclui proporções observadas, estimativas de diferenças em proporções populacionais, erros padrão assintóticos de diferenças populacionais sob hipóteses nulas e alternativas, estatísticas de teste especificadas com probabilidades bilaterais e intervalos de confiança especificados para diferenças em proporções.

### **Exemplo**

## Estatísticas

Agresti-Min, Bonett-Price, Newcombe, Wald, Wald (continuidade corrigida), Binomial Exata, Mid-p Binomial Ajustada, McNemar, McNemar (continuidade corrigida).

## Considerações de dados

### Dados

- São necessárias pelo menos uma variável dependente e uma única variável para identificar os dois grupos a serem comparados.
- A variável de agrupamento pode ser numérica ou uma sequência.

### Suposições

## Obtendo testes de proporções de amostras independentes

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de amostras independentes...**

2. Selecione uma ou mais variáveis de teste quantitativas.

3. Selecione uma única **Variável de agrupamento** que identifica os dois grupos a serem comparados.

4. Opcionalmente, especifique configurações para a **Variável de agrupamento** selecionada.

- Quando **Valores** for selecionado, é possível especificar dois valores numéricos ou de sequência dentro de parênteses para que os valores sejam comparados. Os valores de sequências devem ser colocados entre aspas simples. Os casos com outros valores são ignorados.
- O **Ponto médio** se aplica apenas a variáveis numéricas. Casos no ponto médio ou acima do ponto médio da distribuição da variável de agrupamento são atribuídos ao segundo grupo, casos abaixo do ponto médio são atribuídos ao primeiro grupo.
- O **Ponto de corte** se aplica apenas às variáveis numéricas e permite a especificação com parênteses de um valor numérico único. Casos no ponto de corte ou acima do ponto de corte na variável de agrupamento são atribuídos ao segundo grupo, casos abaixo do ponto de corte são atribuídos ao primeiro grupo.

5. Opcionalmente, é possível:

- Selecione as configurações de critérios de sucesso na seção **Definir sucesso:**

#### Último valor

O último ou o maior valor entre os valores distintos classificados nos dados é usado. Isso se aplica a variáveis numéricas ou de sequência. Essa é a configuração padrão.

#### Primeiro valor

O primeiro ou o menor valor entre os valores distintos classificados nos dados é usado. Isso se aplica a variáveis numéricas ou de sequência.

#### Valor(es)

Um ou mais valores específicos entre parênteses Vários valores devem ser separados por espaços. Isso se aplica a variáveis numéricas ou de sequência. Os valores das variáveis de sequência devem ser colocados entre aspas simples.

#### Ponto médio

Os valores no ou acima do meio do intervalo de valores observados nos dados. Isso se aplica apenas a dados numéricos.

#### Ponto de Corte

Valores no ou acima de um valor especificado. Isso se aplica apenas a dados numéricos.

- Clique em **Intervalos de confiança...** para especificar quais tipos de intervalos de confiança são exibidos ou para suprimir todos os intervalos de confiança.
- Clique em **Testes...** para especificar quais tipos de estatísticas de teste são exibidas ou para suprimir todos os testes.



- Clique em **Valores omissos...** para controlar o tratamento de dados omissos.
- Clique em **Autoinicialização...** para derivar estimativas robustas de erros padrão e intervalos de confiança para estimativas como a média, mediana, proporção, razão de chances, coeficiente de correlação ou coeficiente de regressão.

6. Clique em **OK**.

## Proporções de amostras independentes: intervalos de confiança

A caixa de diálogo Intervalos de confiança fornece opções para especificar o nível de cobertura e para selecionar quais tipos de intervalos de confiança são exibidos.

### Nível de Cobertura

Especifica a porcentagem do intervalo de confiança. Um valor numérico no intervalo (0,100) deve ser especificado. 95 é a configuração padrão.

### Tipo(s) de intervalo

Fornece opções para especificar quais tipos de intervalos de confiança são exibidos. As opções disponíveis incluem:

- Agresti-Caffo
- Brown-Li-Jeffreys
- Hauck-Anderson
- Newcombe
- Newcombe (continuidade corrigida)
- Wald
- Wald (Continuidade corrigida)

## Especificando intervalos de confiança para proporções de amostras independentes

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de amostras independentes...**

2. Clique em **Intervalos de confiança** para especificar quais tipos de intervalos de confiança são exibidos ou para suprimir todos os intervalos de confiança.

## Proporções de amostras independentes: testes

O diálogo Testes fornece opções para especificar quais tipos de estatísticas de teste são exibidos.

### Todos

Todas as estatísticas de teste são exibidas na saída.

### Nenhum

Nenhuma estatística de teste é exibida na saída.

### Hauck-Anderson

Exibe a estatística de teste Hauck-Anderson Z.

### Wald

Exibe a estatística de teste Wald Z.

### Wald (Continuidade corrigida)

Exibe a estatística de teste Wald Z corrigida pela continuidade.

### Wald H<sub>0</sub>

Exibe a estatística de teste Wald Z usando estimativas de variância sob H<sub>0</sub>.

### Wald H<sub>0</sub> (continuidade corrigida)

Exibe a estatística de teste Wald Z corrigida de continuidade usando estimativas de variância sob H<sub>0</sub>.

## Obtendo testes de proporções de amostras independentes

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de amostras independentes...**

2. No diálogo Proporções de amostras independentes, clique em **Testes**.

3. Selecione um ou mais dos testes disponíveis.

## Proporções de amostras independentes: valores omissos

O diálogo Valores omissos fornece opções para lidar com valores omissos.

### Escopo de dados ausente

#### Excluir análise de casos por análise

Indica a inclusão de todos os casos com dados suficientes sobre as variáveis usadas em cada análise particular. Essa é a configuração padrão.

#### Excluir casos de listwise

Indica a inclusão de todos os casos com dados suficientes sobre todas as variáveis usadas em todas as análises.

### Valores ausentes do usuário

**Excluir** trata os valores omissos do usuário como omissos. **Incluir** ignora designações de valor omissos do usuário e trata os valores omissos do usuário como válidos.

## Definindo configurações de valor omissos para Proporções de amostras independentes

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Proporções de amostras independentes...**

2. No diálogo Proporções de amostras independentes, clique em **Valores omissos**.

3. Selecione as configurações de valores omissos desejadas.

## Testes t

---

### Teste t

Três tipos de testes  $t$  estão disponíveis:

**Teste t de amostras independentes (teste T de duas amostras).** Compara o meio de uma variável para dois grupos de casos. Estatísticas descritivas para cada grupo e o teste de Levene para igualdade de variações são fornecidos, bem como os valores de  $t$  de variação igual e desigual e um intervalo de confiança de 95% para a diferença em média.

**Teste t de amostras pairwise teste t dependente).** Compara a média de duas variáveis para um único grupo. Este teste também é destinado a pares correspondidos ou designs de estudo de caso-controle. A saída inclui estatísticas descritivas para as variáveis de teste, a correlação entre as variáveis, as estatísticas descritivas para diferenças pairwise, o teste  $t$  e um intervalo de confiança de 95%.

**Teste t de uma amostra.** Compara a média de uma variável com um valor conhecido ou hipotético. Estatísticas descritivas para as variáveis de teste são exibidas junto com o teste  $t$ . Um intervalo de confiança de 95% para a diferença entre a média da variável do teste e o valor de teste hipotético é parte da saída padrão.

### Teste T de amostras independentes

O procedimento de teste T de amostras independentes compara médias para dois grupos de casos e automatiza o cálculo de tamanho do efeito do teste  $t$ . Idealmente, para este teste, os sujeitos devem ser designados aleatoriamente para dois grupos, de forma que qualquer diferença na resposta será devido

ao tratamento (ou à falta de tratamento) e não a outros fatores. Isso não se aplica se você comparar a renda média de homens e mulheres. Uma pessoa não é designada aleatoriamente para ser homem ou mulher. Nessas situações, é necessário assegurar-se de que as diferenças nos outros fatores não venham mascarar ou aprimorar uma diferença significativa nas médias. As diferenças na renda média podem ser influenciadas por fatores como educação (e não apenas pelo sexo).

### Exemplo

Pacientes com pressão alta são aleatoriamente designados para um grupo placebo e um grupo de tratamento. Os sujeitos do placebo recebem uma pílula inativa e os sujeitos do tratamento recebem uma nova droga que se espera que reduza a pressão arterial. Após dois meses de tratamento, o teste  $t$  de duas amostras é utilizado para comparar a pressão arterial média do grupo placebo e do grupo de tratamento. Cada paciente é avaliado uma vez e pertence a um grupo.

### Estatísticas

Para cada variável: tamanho da amostra, média, desvio padrão, erro padrão da média e a estimação do tamanho do efeito para o teste  $t$ . Para a diferença nas médias: média, erro padrão e intervalo de confiança (é possível especificar o nível de confiança). Testes: teste de Levene de igualdade de variâncias e testes  $t$  de variâncias agrupadas e variâncias separadas para igualdade de médias.

## Considerações de dados

### Dados

Os valores da variável quantitativa de interesse estão em uma única coluna no arquivo de dados. O procedimento usa uma variável de agrupamento com dois valores para separar os casos em dois grupos. A variável de agrupamento pode ser numérica (valores como 1 e 2 ou 6,25 e 12,5) ou uma sequência curta (como *sim* e *não*). Como alternativa, é possível utilizar uma variável quantitativa, como *idade*, para dividir os casos em dois grupos especificando um ponto de corte (o ponto de corte 21 divide *idade* em um grupo sub-21 e um grupo 21 e acima).

### Suposições

Para o teste  $t$  de variância igual, as observações devem ser amostras aleatórias independentes das distribuições normais com a mesma variância da população. Para o teste  $t$  de variância desigual, as observações devem ser amostras aleatórias independentes das distribuições normais. O teste  $t$  de duas amostras é bastante robusto para partidas da normalidade. Ao verificar as distribuições graficamente, verifique se elas são simétricas e não possuem valores discrepantes.

## Obtendo um teste T de amostras independentes

1. A partir dos menus, escolha:

**Analisar > Comparar Médias > Teste T de Amostras Independentes ...**

2. Selecione uma ou mais variáveis de teste quantitativas. Um teste  $t$  separado é calculado para cada variável.

3. Selecione uma variável de agrupamento única e, em seguida, clique em **Definir Grupos** para especificar dois códigos para os grupos que deseja comparar.

4. Opcionalmente, é possível:

- Selecione **Estimar tamanhos do efeito** para controlar a estimação do tamanho do efeito do teste  $t$ .
- Clique em **Opções** para controlar o tratamento de dados omissos e o nível do intervalo de confiança.
- Clique em **Autoinicialização** para derivar estimativas robustas de erros padrão e intervalos de confiança para estimativas, como a média, mediana, proporção, razão de chances, coeficiente de correlação ou coeficiente de regressão.

## Definir grupos de teste T de amostras independentes

Para variáveis de agrupamento numéricas, defina os dois grupos para o teste  $t$  especificando dois valores ou um ponto de corte:

- **Utilizar valores especificados.** Insira um valor para o Grupo 1 e outro valor para o Grupo 2. Casos com quaisquer outros valores são excluídos da análise. Os números não precisam ser inteiros (por exemplo, 6,25 e 12,5 são válidos).
- **Ponto de Corte.** Insira um número que divide os valores da variável de agrupamento em dois conjuntos. Todos os casos com valores menores que o ponto de corte formam um grupo e os casos com valores maiores ou iguais ao ponto de corte formam o outro grupo.

Para variáveis de agrupamento de sequências de caracteres, insira uma sequência de caracteres para o Grupo 1 e outro valor para o Grupo 2, como *yes* e *no*. Casos com outras sequências de caracteres são excluídos da análise.

## Opções de teste T de amostras independentes

**Intervalo de Confiança.** Por padrão, um intervalo de confiança de 95% para a diferença nas médias é exibido. Insira um valor entre 1 e 99 para solicitar um nível de confiança diferente.

**Valores omissos.** Quando você testa várias variáveis e há dados ausentes para uma ou mais variáveis, é possível informar ao procedimento quais casos incluir (ou excluir).

- **Excluir análise de casos por análise.** Cada teste *t* utiliza todos os casos que possuem dados válidos para as variáveis testadas. Os tamanhos das amostras podem variar de teste para teste.
- **Excluir listwise dos casos.** Cada teste *t* utiliza apenas os casos que possuem dados válidos para todas as variáveis que são utilizadas nos testes *t* solicitados. O tamanho da amostra é constante entre os testes.

## Teste-T de amostras em pares

O procedimento do Teste T de Amostras Pareladas compara as médias de duas variáveis para um único grupo. O procedimento calcula as diferenças entre os valores das duas variáveis para cada caso e testa se a média difere de 0. O procedimento também automatiza o cálculo do tamanho do efeito do teste *t*.

### Exemplo

Em um estudo sobre pressão alta, todos os pacientes são medidos no início do estudo, recebem um tratamento e são medidos novamente. Assim, cada sujeito possui duas medidas, geralmente chamadas de medidas *antes* e *após*. Um desenho alternativo para o qual este teste é usado é um estudo de pares pareados ou caso-controle, no qual cada registro no arquivo de dados contém a resposta para o paciente e também para seu sujeito controle pareado. Em um estudo de pressão arterial, pacientes e controles podem ser pareados por idade (um paciente de 75 anos com um membro do grupo controle de 75 anos).

### Estatísticas

Para cada variável: média, tamanho da amostra, desvio padrão e erro padrão da média. Para cada par de variáveis: correlação, diferença média em médias, teste *t*, intervalo de confiança para diferença média (é possível especificar o nível de confiança) e a estimação do tamanho do efeito para o teste *t*. Desvio padrão e erro padrão da diferença média.

## Considerações de dados

### Dados

Para cada teste emparelhado, especifique duas variáveis quantitativas (nível de medição de intervalo ou nível de medição de razão). Para um estudo de correspondência de pares e de caso de controle, a resposta para cada sujeito de teste e seu sujeito de controle correspondente deve estar no mesmo caso no arquivo de dados.

### Suposições

Observações para cada par devem ser feitas nas mesmas condições. As diferenças de médias devem ser normalmente distribuídas. As variâncias de cada variável podem ser iguais ou desiguais.

## Obtendo um teste T de amostras pareadas

1. A partir dos menus, escolha:

## Analisar > Comparar Médias > Teste T de Amostras Emparelhadas ...

2. Selecione um ou mais pares de variáveis.
3. Opcionalmente, mude/selecione uma opção **Estimar tamanhos do efeito**. As configurações controlam como o padronizador é calculado na estimação do  $d$  de Cohen e da correção de Hedges para cada par de variáveis.

### Desvio padrão da diferença

O denominador usado para estimar o tamanho do efeito. O  $d$  de Cohen usa o desvio padrão de amostra da diferença média. A correção do Hedges utiliza o desvio padrão amostral da diferença média ajustada por um fator de correção.

### Desvio padrão corrigido da diferença

O denominador usado para estimar o tamanho do efeito. O  $d$  de Cohen usa o desvio padrão de amostra da diferença média ajustada pela correlação entre medidas. A correção do Hedges utiliza o desvio padrão amostral da diferença média ajustada pela correlação entre as medidas, acrescido de um fator de correção.

### Média de variâncias

O denominador usado para estimar o tamanho do efeito. O  $d$  de Cohen usa a raiz quadrada da variância média de medidas. A correção do Hedges utiliza a raiz quadrada da variância média das medidas, mais um fator de correção.

4. Opcionalmente, é possível:

- Selecione **Estimar tamanhos do efeito** para controlar a estimação do tamanho do efeito do teste  $t$ . Quando a configuração for selecionada, será possível controlar melhor como o padronizador é calculado na estimação do  $d$  de Cohen e da correção de Hedges para cada par de variáveis.
- Clique em **Opções** para controlar o tratamento de dados omissos e o nível do intervalo de confiança.
- Clique em **Autoinicialização** para derivar estimativas robustas de erros padrão e intervalos de confiança para estimativas, como a média, mediana, proporção, razão de chances, coeficiente de correlação ou coeficiente de regressão.

## Opções de Teste T de Amostras Emparelhadas

**Intervalo de Confiança.** Por padrão, um intervalo de confiança de 95% para a diferença nas médias é exibido. Insira um valor entre 1 e 99 para solicitar um nível de confiança diferente.

**Valores omissos.** Quando você testa várias variáveis e faltam dados para uma ou mais variáveis, você pode informar ao procedimento quais casos incluir (ou excluir):

- **Excluir análise de casos por análise.** Cada teste  $t$  utiliza todos os casos que possuem dados válidos para o par de variáveis testado. Os tamanhos das amostras podem variar de teste para teste.
- **Excluir listwise dos casos.** Cada teste  $t$  utiliza apenas os casos que possuem dados válidos para todos os pares de variáveis testados. O tamanho da amostra é constante entre os testes.

## Recursos adicionais do comando T TEST

O idioma da sintaxe de comando também permite:

- Produza testes  $t$  de uma amostra e amostras independentes executando um único comando.
- Testar uma variável com relação a cada variável em uma lista em um teste  $t$  emparelhado (com o subcomando PAIRS).
- Controlar a estimação do tamanho do efeito do teste  $t$  (com o subcomando ES).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Teste-T de uma amostra

O procedimento de teste T de uma amostra testa se a média de uma variável única difere de uma constante especificada e automatiza o cálculo de tamanho do efeito do teste  $t$ .

## Exemplos

Um pesquisador pode querer testar se a pontuação média de QI para um grupo de alunos é diferente de 100. Ou um fabricante de cereais pode pegar uma amostra de caixas da linha de produção e verificar se o peso médio das amostras difere de 1,3 quilo com um nível de confiança de 95%.

## Estatísticas

Para cada variável de teste: média, desvio padrão, erro padrão da média e a estimação do tamanho do efeito para o teste  $t$ . A diferença média entre cada valor de dados e o valor de teste hipotético, um teste  $t$  que confirma que essa diferença é 0, e um intervalo de confiança para esta diferença (é possível especificar o nível de confiança).

## Considerações de dados

### Dados

Para testar os valores de uma variável quantitativa com relação a um valor de teste hipotético, escolha uma variável quantitativa e insira um valor de teste hipotético.

### Suposições

Este teste assume que os dados são normalmente distribuídos; no entanto, este teste é bastante robusto para desvios da normalidade.

## Obtendo um teste T de uma amostra

1. A partir dos menus, escolha:

**Analisar > Comparar médias > Teste T de uma amostra...**

2. Selecione uma ou mais variáveis a serem testadas em relação ao mesmo valor hipotético.

3. Insira um valor de teste numérico com o qual cada média de amostra é comparada.

4. Opcionalmente, é possível:

- Selecione **Estimar tamanhos do efeito** para controlar a estimação do tamanho do efeito do teste  $t$ .
- Clique em **Opções** para controlar o tratamento de dados omissos e o nível do intervalo de confiança.

## Opções de teste T de uma amostra

**Intervalo de Confiança.** Por padrão, um intervalo de confiança de 95% para a diferença entre a média e o valor de teste hipotético é exibido. Insira um valor entre 1 e 99 para solicitar um nível de confiança diferente.

**Valores omissos.** Quando você testa várias variáveis e faltam dados para uma ou mais variáveis, você pode informar ao procedimento quais casos incluir (ou excluir).

- **Excluir análise de casos por análise.** Cada teste  $t$  utiliza todos os casos que possuem dados válidos para a variável testada. Os tamanhos das amostras podem variar de teste para teste.
- **Excluir listwise dos casos.** Cada teste  $t$  utiliza apenas os casos que possuem dados válidos para todas as variáveis que são utilizadas em qualquer um dos testes  $t$  solicitados. O tamanho da amostra é constante entre os testes.

## Recursos adicionais do comando T TEST

O idioma da sintaxe de comando também permite:

- Produza testes  $t$  de uma amostra e amostras independentes executando um único comando.
- Testar uma variável com relação a cada variável em uma lista em um teste  $t$  emparelhado (com o subcomando PAIRS).
- Controlar a estimação do tamanho do efeito do teste  $t$  (com o subcomando ES).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Recursos adicionais do comando T TEST

O idioma da sintaxe de comando também permite:

- Produza testes  $t$  de uma amostra e amostras independentes executando um único comando.
- Testar uma variável com relação a cada variável em uma lista em um teste  $t$  emparelhado (com o subcomando PAIRS).
- Controlar a estimação do tamanho do efeito do teste  $t$  (com o subcomando ES).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## ANOVA de um fator

---

O procedimento ANOVA de um fator produz uma análise de variância de um fator para uma variável dependente quantitativa por uma variável de fator único (independente) e estima o tamanho do efeito em ANOVA de um fator. A análise de variância é utilizada para testar a hipótese de que várias médias são iguais. Esta técnica é uma extensão do teste  $t$  de duas amostras.

Além de determinar que existem diferenças entre as médias, talvez você queira saber quais médias diferem. Há dois tipos de testes para comparar médias: contrastes a priori e testes post hoc. Contrastes são testes configurados *antes de* executar o experimento, e testes post hoc são executados *após* o experimento ter sido conduzido. Também é possível testar tendências entre as categorias.

### Exemplo

Doces como sonhos de padarias absorvem várias quantias de gordura quando são preparados. Um experimento é montado envolvendo três tipos de gordura: óleo de amendoim, óleo de milho e banha. O óleo de amendoim e o óleo de milho são gorduras insaturadas e a banha é uma gordura saturada. Além de determinar se a quantidade de gordura absorvida depende do tipo de gordura usada, você pode configurar um contraste a priori para determinar se a quantidade de absorção de gordura difere para gorduras saturadas e insaturadas.

### Estatísticas

Para cada grupo: número de casos, média, desvio padrão, erro padrão da média, mínimo, máximo, intervalo de confiança de 95% para a média e a estimativa do tamanho do efeito para uma ANOVA de um fator. Testes de homogeneidade de variâncias de Levene, tabela de análise de variância e testes robustos de igualdade de médias para cada variável dependente, contrastes especificados a priori pelo usuário e testes de intervalo a posteriori e várias comparações: Bonferroni, Sidak, diferença honestamente significativa de Tukey, GT2 de Hochberg, Gabriel, Dunnett, teste  $F$  de Ryan-Einot-Gabriel-Welsch (R-E-G-W  $F$ ), teste de amplitude de Ryan-Einot-Gabriel-Welsch (R-E-G-W  $Q$ ), T2 de Tamhane, T3 de Dunnett, Games-Howell,  $C$  de Dunnett, teste de amplitude variada de Duncan, Student-Newman-Keuls (S-N-K),  $b$  de Tukey, Waller-Duncan, Scheffé, e diferença menos significativa.

## Considerações de dados

### Dados

A variável dependente deve ser quantitativa (nível de intervalo de medição).

### Suposições

Cada grupo é uma amostra aleatória independente de uma população normal. A análise de variância é robusta para partidas da normalidade, embora os dados devam ser simétricos. Os grupos devem vir de populações com variâncias iguais. Para testar essa suposição, use o teste de homogeneidade de variância de Levene.

## Obtendo uma análise de variância para um fator

1. A partir dos menus, escolha:

**Analisar > Comparar médias > ANOVA de um fator...**

2. Selecione uma ou mais variáveis dependentes.

3. Selecione uma única variável de fator independente.

Opcionalmente, é possível:

- Selecione **Estimar tamanho do efeito para testes gerais** para controlar o cálculo do tamanho do efeito para o teste geral. Quando selecionado, a tabela “Tamanhos de efeito ANOVA” é exibida na saída.
- Clique em **Contrastes** para particionar as somas de quadrados entre grupos para componentes de tendência ou especificar contrastes a priori.
- Clique em **Posteriori** para usar testes de intervalo posteriori e comparações múltiplas pairwise para determinar quais médias diferem.
- Clique em **Opções** para controlar o tratamento de dados omissos e o nível do intervalo de confiança.
- Clique em **Autoinicialização** para derivar estimativas robustas de erros padrão e intervalos de confiança para estimativas, como a média, mediana, proporção, razão de chances, coeficiente de correlação ou coeficiente de regressão.

## Contrastes ANOVA de um fator

É possível particionar as somas dos quadrados entre grupos em componentes de tendência ou especificar contrastes a priori.

### Polinomial

Particiona as somas dos quadrados entre grupos em componentes de tendência. É possível testar para uma tendência da variável dependente entre os níveis ordenados da variável de fator. Por exemplo, você pode testar uma tendência linear (aumentando ou diminuindo) no salário nos níveis ordenados do grau mais alto obtido.

- **Grau.** É possível escolher um polinomial de 1º, 2º, 3º, 4º ou 5º grau.

### Coeficientes

Contrastes a priori especificados pelo usuário a serem testados pela estatística *t*. Insira um coeficiente para cada grupo (categoria) da variável de fator e clique em **Incluir** após cada entrada. Cada novo valor é incluído na parte inferior da lista de coeficientes. Para especificar conjuntos de contrastes adicionais, clique em **Avançar**. Utilize **Avançar** e **Anterior** para mover entre os conjuntos de contrastes.

### Estimar o tamanho do efeito para contrastes

Controla o cálculo do tamanho do efeito para o teste geral. Quando esta configuração é ativada, pelo menos uma das opções a seguir deve ser selecionada para calcular os tamanhos de efeito. Essa configuração é ativada quando pelo menos um contraste é especificado e resulta em uma tabela de tamanhos de efeito ANOVA na saída.

#### Usar desvio padrão agrupado para todos os grupos como o padronizador

Usa o desvio padrão agrupado para todos os grupos como o padronizador para estimar o tamanho do efeito na estimação do tamanho do efeito. Esta é a configuração padrão e está disponível quando **Estimar tamanho do efeito para contrastes** é selecionado.

#### Usar desvio padrão agrupado para os grupos envolvidos no contraste como o padronizador

Usa o desvio padrão agrupado para os grupos envolvidos no contraste como o padronizador. A configuração está disponível quando **Estimar tamanho do efeito para contrastes** é selecionado.

A ordem dos coeficientes é importante porque corresponde à ordem crescente dos valores de categoria da variável fator. O primeiro coeficiente da lista corresponde ao valor mais baixo do grupo da variável de fator e o último coeficiente corresponde ao valor mais alto. Por exemplo, se houver seis categorias da variável de fator, os coeficientes -1, 0, 0, 0, 0,5 e 0,5 contrastam o primeiro grupo com o quinto e sexto grupos. Para a maioria dos aplicativos, os coeficientes devem ter soma 0. Os conjuntos que não somam 0 também podem ser usados, mas uma mensagem de aviso é exibida.

## Testes posteriori de ANOVA de um fator

Depois de determinar que existem diferenças entre as médias, testes de intervalo posteriori e comparações múltiplas aos pares podem determinar quais médias diferem. Os testes de alcance identificam subconjuntos homogêneos de médias que não são diferentes entre si. As múltiplas



comparações de pares testam a diferença entre cada par de médias e produzem uma matriz onde os asteriscos indicam médias de grupo significativamente diferentes em um nível alfa de 0,05.

## Variâncias iguais presumidas

O teste da diferença significativa honesta de Tukey, GT2 de Hochberg, Gabriel e Scheffé são testes de comparação múltipla e também testes de intervalo. Outros testes de intervalo disponíveis são *b* de Tukey, *S-N-K* (Student-Newman-Keuls), Duncan, *F* de R-E-G-W (teste *F* de Ryan-Einot-Gabriel-Welsch), *Q* de R-E-G-W (teste de intervalo de Ryan-Einot-Gabriel-Welsch) e Waller-Duncan. Os testes de múltiplas comparações disponíveis são Bonferroni, teste de diferença honestamente significativa de Tukey, Sidak, Gabriel, Hochberg, Dunnett, Scheffé e LSD (diferença menos significativa).

- *LSD*. Usa os testes *t* para executar todas as comparações pairwise entre as médias do grupo Nenhum ajustamento é feito na taxa de erros para comparações múltiplas.
- *Bonferroni*. Utiliza testes *t* para realizar comparações pairwise entre os meios de grupo, mas controla a taxa de erro geral configurando a taxa de erro para cada teste à taxa de erro de experimentwise dividida pelo número total de testes. Portanto, o nível de significância observado é ajustado para o fato de que diversas comparações estão sendo feitas.
- *Sidak*. Teste de comparação múltipla entre pares com base em uma estatística *t* O Sidak ajusta o nível de significância para comparações múltiplas e fornece limites mais apertados que Bonferroni.
- *Scheffe*. Executa comparações pairwise conjuntas simultâneas para todas as combinações de médias pairwise possíveis. Utiliza a distribuição *F* de amostragem. Pode ser utilizado para examinar todas as combinações lineares possíveis das médias de grupo, não apenas comparações pairwise.
- *R-E-G-W F*. Procedimento stepdown múltiplo de Ryan-Einot-Gabriel-Welsch com base em um teste *F*.
- *R-E-G-W Q*. Procedimento stepdown múltiplo de Ryan-Einot-Gabriel-Welsch com base em um intervalo estudentizado.
- *S-N-K*. Faz todas as comparações entre pares entre as médias usando a distribuição de intervalo estudentizado Com tamanhos de amostra iguais, ele também compara pares de médias em subconjuntos homogêneos, utilizando um procedimento stepwise. As médias são ordenadas da mais alta para a mais baixa, e as diferenças extremas são testadas em primeiro lugar.
- *Tukey*. Usa a estatística de intervalo estudentizado para fazer todas as comparações pairwise entre os grupos Configura a taxa de erros entre os experimentos na taxa de erros da coleção de todas as comparações pairwise.
- *b de Tukey*. Usa a distribuição do intervalo estudentizado para fazer comparações entre pares entre os grupos O valor crítico é a média do valor correspondente para o teste da diferença significativa honesta de Tukey e o Student-Newman-Keuls.
- *Duncan*. Faz comparações pairwise usando uma ordem de comparações stepwise idêntica à ordem usada pelo teste Student-Newman-Keuls, mas configura um nível de proteção para a taxa de erro para a coleção de testes, em vez de uma taxa de erros para testes individuais. Utiliza a amplitude estatística estudentizada.
- *GT2 de Hochberg*. Teste de comparação e intervalo múltiplos que usa o módulo máximo estudentizado. Semelhante ao teste da diferença significativa honesta de Tukey.
- *Gabriel*. Teste de comparação entre pares que usou o módulo máximo estudentizado e geralmente é mais poderoso do que GT2 de Hochberg quando os tamanhos de célula são desiguais. O teste de Gabriel pode ser liberal quando os tamanhos de células variam grandiosamente.
- *Waller-Duncan*. Teste de comparação múltipla com base em uma estatística *t* que utiliza abordagem Bayesiana.
- *Dunnett*. Teste de múltipla comparação pairwise *t* que compara um conjunto de tratamentos contra uma média de controle único. A última categoria é a categoria de controle padrão. Alternativamente, é possível escolher a primeira categoria. **Bilaterais** testa que a média em qualquer nível (exceto na categoria de controle) do fator não é igual à da categoria de controle. **O controle <** testa se a média em qualquer nível do fator é menor que a da categoria de controle. **O controle >** testa se a média em qualquer nível do fator é maior que a da categoria controle.

## Variâncias iguais não presumidas

Testes de comparação múltipla não consideram que variâncias iguais sejam T2 de Tamhane, T3 de Dunnett, Games-Howell e C de Dunnett.

- *T2 de Tamhane*. Teste de comparação entre pares conservador baseado em um teste t. Este teste é apropriado quando as variâncias são desiguais.
- *T3 de Dunnett*. Teste de comparação entre pares com base no módulo máximo estudentizado Este teste é apropriado quando as variâncias são desiguais.
- *Games-Howell*. Teste de comparação entre pares que às vezes é liberal. Este teste é apropriado quando as variâncias são desiguais.
- *C de Dunnett*. Teste de comparação entre pares com base no intervalo estudado. Este teste é apropriado quando as variâncias são desiguais.

**Nota:** Você pode achar mais fácil interpretar a saída de testes a posteriori se desmarcar **Ocultar linhas e colunas vazias** na caixa de diálogo Propriedades da tabela (em uma tabela dinâmica ativada, escolha **Propriedades da tabela** do menu Formatar).

## Teste de hipótese nula

Especifica como o nível de significância (alpha) é manipulado para o teste a posteriori.

### Usar o mesmo nível de significância (alpha) que as configurações em Opções

Quando selecionado, usa a mesma configuração especificada na caixa de diálogo Opções.

### Especifique o nível de significância (alpha) para o teste a posteriori

Quando selecionada, é possível especificar o nível de significância (alpha) no campo **Nível**.

## Obtendo Testes Post Hoc para ANOVA de um fator

## Opções de ANOVA unidirecional

### Estatísticas

Escolha um ou mais dos seguintes:

#### Descritivo

Calcula o número de casos, média, desvio padrão, erro padrão da média, mínimo, máximo e intervalos de confiança de 95% para cada variável dependente de cada grupo.

#### Efeitos fixos e aleatórios

Exibe o desvio padrão, erro padrão e intervalo de confiança de 95% para o modelo de efeitos fixos e o erro padrão, intervalo de confiança de 95% e estimativa de variância entre componentes para o modelo de efeitos aleatórios.

#### Teste de homogeneidade de variância

Calcula a estatística de Levene para testar a igualdade de variâncias de grupos. Esse teste não depende da suposição de normalidade.

#### Brown-Forsythe

Calcula a estatística Brown-Forsythe para testar a igualdade das médias do grupo. Essa estatística é preferível sobre a estatística de  $F$  quando a suposição de variâncias iguais não é mantida.

#### Welch

Calcula a estatística de Welch a ser testada para a igualdade de médias de grupos. Essa estatística é preferível sobre a estatística de  $F$  quando a suposição de variâncias iguais não é mantida.

### Valores omissos

Controla o tratamento de valores ausentes.

#### Excluir análise de casos por análise

Um caso com um valor ausente para a variável dependente ou fator para uma determinada análise não é usado nessa análise. Além disso, um caso fora do intervalo especificado para a variável de fator não é usado.

### **Excluir casos de listwise**

Casos com valores ausentes para a variável de fator ou para qualquer variável dependente incluída na lista de dependentes na caixa de diálogo principal são excluídos de todas as análises. Se você não especificou várias variáveis dependentes, isso não tem efeito.

### **Intervalo de confiança**

Por padrão, um intervalo de confiança de 95% para a diferença entre a média e o valor de teste hipotético é exibido. Insira um valor entre 1 e 99 para solicitar um nível de confiança diferente.

### **Gráfico de médias**

Exibe um gráfico que plota as médias do subgrupo (as médias para cada grupo definido pelos valores da variável de fator).

## **Especificando Opções para ANOVA de um fator**

### **Recursos adicionais do comando ONEWAY**

O idioma da sintaxe de comando também permite:

- Obter estatísticas de efeitos fixo e aleatório. Desvio padrão, erro padrão da média e intervalos de confiança de 95% para o modelo de efeitos fixos. Erro padrão, intervalos de confiança de 95% e a estimativa de variância entre componentes para modelo de efeitos aleatórios (utilizando STATISTICS=EFFECTS).
- Especificar níveis alfa para a diferença menos significativa, Bonferroni, Duncan, e testes de comparação múltipla de Scheffé (com o subcomando RANGES).
- Gravar uma matriz de médias, desvios padrão e frequências ou ler uma matriz de médias, frequências, variâncias agrupadas e graus de liberdade para as variâncias agrupadas. Essas matrizes podem ser utilizadas no lugar de dados brutos para obter uma análise de variância unidirecional (com o subcomando MATRIX).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## **Análise GLM Univariate**

---

O procedimento GLM Univariate fornece análise de regressão e análise de variância para uma variável dependente por um ou mais fatores e/ou variáveis. As variáveis de fator dividem a população em grupos. Usando este procedimento do Modelo Linear Geral, é possível testar hipóteses nulas sobre os efeitos de outras variáveis nas médias de vários agrupamentos de uma única variável dependente. É possível investigar as interações entre fatores, bem como os efeitos de fatores individuais, alguns dos quais podem ser aleatórios. Além disso, os efeitos de covariáveis e de interações de covariáveis com fatores podem ser incluídos. Para análise de regressão, as variáveis independentes (preditoras) são especificadas como covariáveis.

Modelos balanceados e não balanceados podem ser testados. Um design será balanceado se cada célula no modelo contiver o mesmo número de casos. Além de testar hipóteses, o GLM Univariate produz estimativas de parâmetros.

Contrastes a priori normalmente usados estão disponíveis para executar teste de hipótese. Além disso, após um teste  $F$  geral ter mostrado significância, será possível utilizar testes post hoc para avaliar diferenças entre médias específicas. As médias marginais estimadas fornecem estimativas dos valores médios previstos para as células no modelo, e os gráficos de perfil (gráficos de interação) dessas médias permitem visualizar facilmente algumas das relações.

Valores residuais, valores preditos, a distância de Cook e valores de ponto de alavanca podem ser salvos como novas variáveis em seu arquivo de dados para verificação de suposições.

A Ponderação de WLS permite especificar uma variável usada para dar às observações ponderações diferentes para uma análise de quadrados mínimos ponderados (WLS), talvez para compensar a precisão diferente de medição.

**Exemplo.** Os dados são reunidos para corredores individuais na maratona de Chicago por vários anos. O tempo em que cada corredor termina é a variável dependente. Outros fatores incluem clima (frio, agradável ou quente), número de meses de treinamento, número de maratonas anteriores e sexo. A idade é considerada uma covariável. Você pode achar que o gênero é um efeito significativo e que a interação do gênero com o clima é significativa.

**Métodos.** Somas dos quadrados do Tipo I, Tipo II, Tipo III e Tipo IV podem ser utilizadas para avaliar diferentes hipóteses. O Tipo III é o padrão.

**Estatísticas.** Testes de intervalo e comparações múltiplas autoteste inicial posteriori: diferença menos significativa, Bonferroni, Sidak, Scheffé,  $F$  múltiplo de Ryan-Einot-Gabriel-Welsch e amplitude múltipla de Ryan-Einot-Gabriel-Welsch, Aluno-Newman-Keuls, diferença honestamente significativa de Tukey,  $b$  de Tukey, Duncan, GT2 de Hochberg, Gabriel, teste  $t$  de Waller-Duncan, Dunnett (unilateral e bilateral), T2 de Tamhane, T3 de Dunnett, Games-Howell e  $C$  de Dunnett. Estatística descritiva: médias observadas, desvios padrão e contagens para todas as variáveis dependentes em todas as células. Testes de Levene para homogeneidade de variância.

**Gráficos.** Dispersão versus nível, residual e perfil (interação).

Considerações sobre dados do GLM Univariate

**Dados.** A variável dependente é quantitativa. Os fatores são categóricos. Eles podem ter valores numéricos ou valores de string de até oito caracteres. Covariáveis são variáveis quantitativas que estão relacionadas à variável dependente.

**Suposições.** Os dados são uma amostra aleatória de uma população normal; na população, todas as variâncias celulares são as mesmas. A análise de variância é robusta para partidas da normalidade, embora os dados devam ser simétricos. Para verificar suposições, é possível usar testes de homogeneidade de variâncias e gráficos de dispersão versus nível. Também é possível examinar resíduos e gráficos de resíduos.

Para obter tabelas do GLM Univariate

1. A partir dos menus, escolha:

**Analisar > Modelo Linear Geral > Univariada...**

2. Selecione uma variável dependente.

3. Selecione variáveis para Fatores Fixos, Fatores Aleatórios e Covariáveis, conforme apropriado para seus dados.

4. Opcionalmente, é possível usar a ponderação WLS para especificar uma variável de peso para análise de mínimos quadrados ponderados. Se o valor da variável de ponderação for zero, negativo ou omissão, o caso será excluído da análise. Uma variável já utilizada no modelo não pode ser utilizada como uma variável de ponderação.

## Modelo de GLM

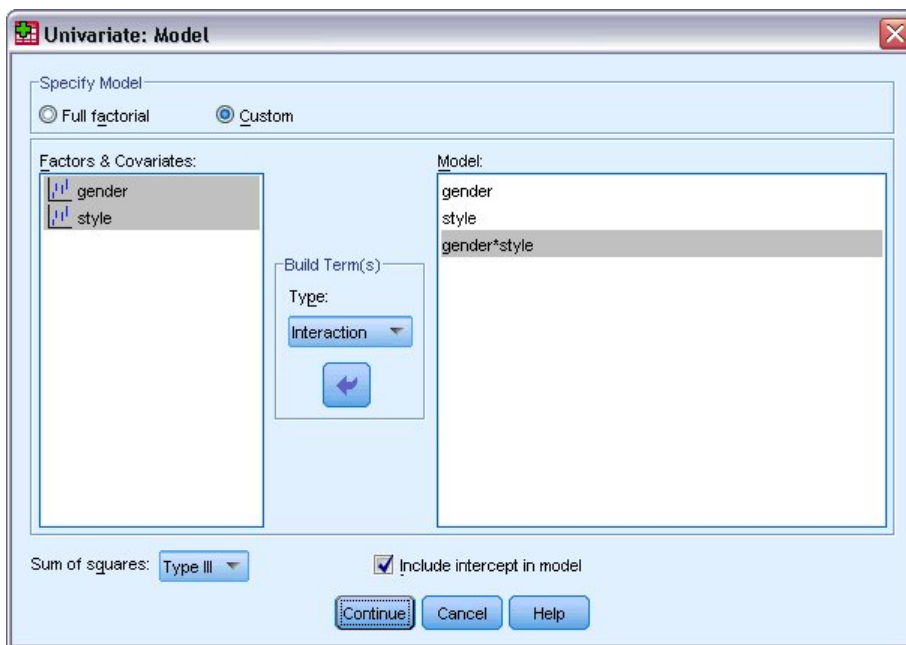


Figura 1. Caixa de diálogo Modelo univariado

**Especificar modelo.** Um modelo fatorial completo contém todos os efeitos principais do fator, todos os efeitos principais das covariáveis e todas as interações fator por fator. Ele não contém interações de covariáveis. Selecione **Customizado** para especificar apenas um subconjunto de interações ou para especificar interações fator por covariável. Deve-se indicar todos os termos a serem incluídos no modelo.

**Fatores e covariáveis.** Os fatores e covariáveis são listados.

**Modelo.** O modelo depende da natureza de seus dados. Depois de selecionar **Customizado**, é possível selecionar os principais efeitos e interações que são de interesse em sua análise.

**Soma dos quadrados.** O método de calcular as somas dos quadrados. Para modelos balanceados ou não balanceados sem células omissas, o método da soma dos quadrados do Tipo III é mais comumente usado.

**Incluir intercepto no modelo.** O intercepto geralmente é incluído no modelo. Se você conseguir presumir que os dados passam pela origem, será possível excluir o intercepto.

### Termos de construção e termos customizados

#### termos de construção

Use essa opção quando desejar incluir termos não aninhados de um determinado tipo (como efeitos principais) para todas as combinações de um conjunto selecionado de fatores e covariáveis.

#### Construir termos customizados

Use esta opção quando desejar incluir termos aninhados ou quando desejar construir explicitamente qualquer variável de termo por variável. A criação de um termo aninhado envolve as etapas a seguir:

### Soma de quadrados

Para o modelo, é possível escolher um tipo de somas dos quadrados. O tipo III é o mais comumente usado e é o padrão.

**Tipo I.** Esse método também é conhecido como decomposição hierárquica do método da soma dos quadrados. Cada termo é ajustado para apenas o termo que o precede no modelo. As somas dos quadrados do tipo I são comumente usadas para:

- Um modelo ANOVA balanceado em que quaisquer efeitos principais serão especificados antes de qualquer efeito de interação de primeira ordem, qualquer efeito de interação de primeira ordem é especificado antes de qualquer efeito de interação de segunda ordem, e assim por diante.
- Um modelo de regressão polinomial em que qualquer termo de ordem inferior é especificado antes de qualquer termo de ordem superior.
- Um modelo puramente aninhado em que o efeito primeiro especificado é aninhado dentro do segundo efeito especificado, o efeito segundo especificado é aninhado dentro do terceiro, e assim por diante. (Esta forma de aninhamento pode ser especificada usando sintaxe.)

**Tipo II.** Este método calcula as somas dos quadrados de um efeito no modelo ajustado para todos os outros efeitos "apropriados". Um efeito apropriado é aquele que corresponde a todos os efeitos que não contêm o efeito que está sendo examinado. O método de soma de quadrados Tipo II é comumente usado para:

- Um modelo ANOVA balanceado.
- Qualquer modelo que possui apenas efeitos de fator principal.
- Qualquer modelo de regressão.
- Um design puramente aninhado. (Esta forma de aninhamento pode ser especificada usando a sintaxe.)

**Tipo III.** O padrão. Este método calcula as somas dos quadrados de um efeito no desenho como as somas dos quadrados, ajustadas para quaisquer outros efeitos que não contenham o efeito e ortogonais a quaisquer efeitos (se houver) que contenham o efeito. A soma dos quadrados do Tipo III possui uma grande vantagem em que eles são invariáveis com relação às frequências de célula contanto que o formato geral de estimabilidade permaneça constante. Portanto, esse tipo de somas de quadrados é geralmente considerado útil para um modelo não balanceado sem células omissas. Em um design fatorial sem células omissas, este método é equivalente à técnica médias dos quadrados ponderados de Yates. O método da soma dos quadrados do Tipo III é comumente utilizado para:

- Todos os modelos listados em Tipo I e Tipo II.
- Qualquer modelo balanceado ou desbalanceado sem células vazias.

**Tipo IV.** Este método é projetado para uma situação em que há células ausentes. Para qualquer efeito  $F$  no design, se  $F$  não está contido em nenhum outro efeito, então Tipo IV = Tipo III = Tipo II. Quando  $F$  está contido em outros efeitos, o Tipo IV distribui os contrastes que estão sendo feitos entre os parâmetros em  $F$  para todos os efeitos de nível superior equitativamente. O método de soma de quadrados Tipo IV é comumente usado para:

- Todos os modelos listados em Tipo I e Tipo II.
- Qualquer modelo balanceado ou modelo não balanceado com células vazias.

## Contrastes GLM

Os contrastes são usados para testar as diferenças entre os níveis de um fator. É possível especificar um contraste para cada fator no modelo (em um modelo de medidas repetidas, para cada fator entre sujeitos). Os contrastes representam as combinações lineares dos parâmetros.

**GLM Univariate.** Os testes de hipótese baseiam-se na hipótese nula  $\mathbf{LB} = 0$ , em que  $\mathbf{L}$  é a matriz de coeficientes de contraste e  $\mathbf{B}$  é o vetor paramétrica. Quando um contraste é especificado, uma matriz  $\mathbf{L}$  é criada. As colunas da matriz  $\mathbf{L}$  correspondem ao fator que corresponde ao contraste. As colunas restantes são ajustadas para que a matriz  $\mathbf{L}$  seja estimável.

A saída inclui uma estatística  $F$  para cada conjunto de contrastes. Também exibidos para as diferenças de contrastes estão os intervalos de confiança simultâneos do tipo Bonferroni com base na distribuição  $t$  de Student.

### Contrastes Disponíveis

Os contrastes disponíveis são desvio, simples, diferença, Helmert, repetido e polinomial. Para contrastes de desvio e contrastes simples, você pode escolher se a categoria de referência é a última ou a primeira categoria.

## Tipos de contraste

**Desvio.** Compara a média de cada nível (exceto uma categoria de referência) com a média de todos os níveis (média global). Os níveis do fator podem estar em qualquer ordem.

**Simples.** Compara a média de cada nível com a média de um nível especificado. Esse tipo de contraste é útil quando há um grupo de controle. É possível escolher a primeira ou a última categoria como referência.

**Diferença.** Compara a média de cada nível (exceto do primeiro) com a média dos níveis anteriores. (Às vezes chamado de contrastes reversos de Helmert.)

**Helmert.** Compara a média de cada nível do fator (exceto do último) com a média dos níveis subsequentes.

**Repetido.** Compara a média de cada nível (exceto do último) com a média dos níveis subsequentes.

**Polinomial.** Compara o efeito linear, o efeito quadrático, o efeito cúbico, e assim por diante. O primeiro grau de liberdade contém o efeito linear em todas as categorias; o segundo grau de liberdade, o efeito quadrático; e assim por diante. Esses contrastes são frequentemente usados para estimar tendências polinomiais.

## Gráficos de perfil GLM

Os gráficos de perfil (gráficos de interação) são úteis para comparar médias marginais em seu modelo. Um gráfico de perfil é um gráfico de linha em que cada ponto indica a média marginal estimada de uma variável dependente (ajustada para quaisquer covariáveis) em um nível de um fator. Os níveis de um segundo fator podem ser usados para fazer linhas separadas. Cada nível em um terceiro fator pode ser usado para criar um gráfico separado. Todos os fatores fixos e aleatórios, se houver, estão disponíveis para gráficos. Para análises multivariadas, são criados gráficos de perfil para cada variável dependente. Em uma análise de medidas repetidas, tanto os fatores entre sujeitos quanto os fatores dentro dos sujeitos podem ser usados em gráficos de perfil. GLM Multivariate e GLM Repeated Measures estão disponíveis apenas se você tiver a opção Estatísticas avançadas instalada.

Um gráfico de perfil de um fator mostra se as médias marginais estimadas estão aumentando ou diminuindo entre os níveis. Para dois ou mais fatores, as linhas paralelas indicam que não há interação entre os fatores, o que significa que você pode investigar os níveis de apenas um fator. Linhas não paralelas indicam uma interação.

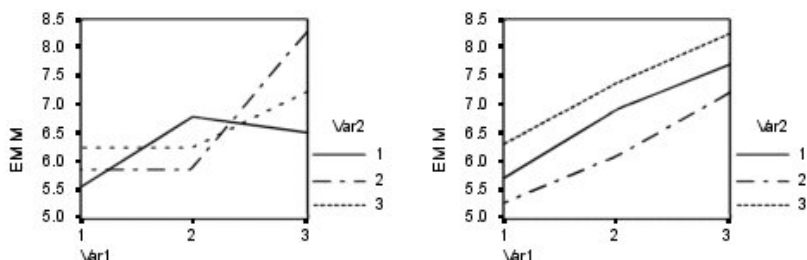


Figura 2. Gráfico não paralelo (esquerdo) e gráfico paralelo (direito)

Depois que um gráfico é especificado selecionando fatores para o eixo horizontal e, opcionalmente, fatores para linhas separadas e gráficos separados, o gráfico deve ser adicionado à lista Gráficos.

### Chart Type

O gráfico pode ser um gráfico de linha ou um gráfico de barras

### Barras de erros

É possível incluir barras de erro que representam o intervalo de confiança ou vários erros padrão. O intervalo de confiança é baseado no nível de significância especificado na caixa de diálogo Opções.

### Incluir linha de referência para média global

Inclui uma linha de referência que representa a média global geral.

## Eixo Y inicia em 0

Para gráficos de linhas com todos os valores positivos ou todos os negativos, força o eixo Y a começar em 0. Os gráficos de barras sempre começam (ou incluem) 0.

## Opções de GLM

Estatísticas de opcionais estão disponíveis a partir dessa caixa de diálogo. As estatísticas são calculadas utilizando um modelo de efeitos fixos.

**Exibição.** Selecione **Estatísticas descritivas** para produzir médias observadas, desvios padrão e contagens para todas as variáveis dependentes em todas as células. **Estimativas de tamanho de efeito** fornece um valor eta quadrado parcial para cada efeito e cada estimativa paramétrica. A estatística eta quadrado descreve a proporção da variabilidade total atribuível a um fator. Selecione **Potência observada** para obter a potência do teste quando a hipótese alternativa é configurada com base no valor observado. Selecione **Estimativas paramétrica** para produzir as estimativas paramétrica, os erros padrão, testes *t*, intervalos de confiança e a potência observada para cada teste. Selecione **Matriz de coeficientes de contraste** para obter a matriz **L**.

Os **Testes de homogeneidade** produzem testes de Levene da homogeneidade da variância para cada variável dependente em todas as combinações de nível dos fatores entre assuntos, somente para fatores entre assuntos. As opções de gráficos de dispersão versus nível e de resíduos são úteis para verificar suposições sobre os dados. Esse item ficará desativado se não houver fatores. Selecione **Gráfico de resíduo** para produzir um gráfico residual observado por predito por padronizado para cada variável dependente. Esses gráficos são úteis para investigar a suposição de variância igual. Selecione **Falta de ajuste** para verificar se o relacionamento entre a variável dependente e as variáveis independentes pode ser descrito adequadamente pelo modelo. **Funções estimáveis gerais** permite construir testes de hipótese customizados com base nas funções estimáveis gerais. As linhas em qualquer matriz de coeficiente de contraste são combinações lineares das funções estimáveis gerais.

Os **Testes de heterocedasticidade** estão disponíveis para testar se a variância dos erros (para cada variável dependente) depende dos valores das variáveis independentes. Para o **Teste de Breusch-Pagan**, **Teste de Breusch-Pagan modificado** e o **Teste F**, é possível especificar o modelo no qual o teste é baseado. Por padrão, o modelo consiste em um termo constante, um termo que é linear nos valores preditos, um termo que é quadrático nos valores preditos e um termo de erro.

As **Estimativas de parâmetro com erros padrão robustos** exibem uma tabela de estimativas de parâmetro, juntamente com erros padrão robustos ou consistentes com heterocedasticidade (HC); além de estatísticas de *t*, valores de significância e intervalos de confiança que usam os erros padrão robustos. Há cinco métodos diferentes disponíveis para a estimativa da matriz de covariâncias robusta.

### HC0

Baseado no assintótico original ou no estimador robusto de amostra grande, empírico ou "sanduíche" da matriz de covariâncias das estimativas de parâmetro. A parte do meio do sanduíche contém resíduos de OLS (quadrados mínimos ordinários) quadrados ou de WSL (quadrados mínimos ponderados) ponderados quadrados.

### HC1

Uma modificação de amostra finita de HC0, multiplicada por  $N/(N-p)$ , em que  $N$  é o tamanho da amostra e  $p$  é o número de parâmetros não redundantes no modelo.

### HC2

Uma modificação de HC0 que envolve a divisão do resíduo quadrado por  $1-h$ , em que  $h$  é a alavanca para o caso.

### HC3

Uma modificação de HC0 que aproxima um estimador jackknife. Os resíduos quadrados são divididos pelo quadrado de  $1-h$ .

### HC4

Uma modificação de HC0 que divide os resíduos quadrados por  $1-h$  a uma potência que varia de acordo com  $h$ ,  $N$  e  $p$ , com um limite superior de 4.



**Nível de Significância.** Talvez você queira ajustar o nível de significância utilizado em testes posteriori e o nível de confiança utilizado para construir intervalos de confiança. O valor especificado é também usado para calcular o poder observado para o teste. Ao especificar um nível de significância, o nível associado dos intervalos de confiança é exibido na caixa de diálogo.

## Recursos adicionais do comando UNIANOVA

O idioma da sintaxe de comando também permite:

- Especificar efeitos aninhados no design (utilizando o subcomando DESIGN).
- Especificar testes de efeitos versus uma combinação linear de efeitos ou um valor (utilizando o subcomando TEST).
- Especificar diversos contrastes (utilizando o subcomando CONTRAST).
- Incluir valores omissos de usuário (utilizando o subcomando MISSING).
- Especificar critérios do ESP (utilizando o subcomando CRITERIA).
- Construir uma matriz **L**, uma matriz **M** ou uma matriz **K** customizada (usando os subcomandos LMATRIX, MMATRIX e KMATRIX).
- Para desvio ou contrastes simples, especifique uma categoria de referência intermediária (utilizando o subcomando CONTRAST).
- Especificar métricas para contrastes polinomiais (utilizando o subcomando CONTRAST).
- Especificar termos de erro para comparações post hoc (utilizando o subcomando POSTHOC).
- Calcular médias marginais estimadas para qualquer fator ou interação entre fatores entre os fatores na lista de fatores (utilizando o subcomando EMMEANS).
- Especificar nomes para variáveis temporárias (utilizando o subcomando SAVE).
- Construir um arquivo de dados de matriz de correlações (utilizando o subcomando OUTFILE).
- Construir um arquivo de dados de matriz que contenha estatísticas da tabela ANOVA entre assuntos (utilizando o subcomando OUTFILE).
- Salvar a matriz de design em um novo arquivo de dados (utilizando o subcomando OUTFILE).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Comparações a posteriori de GLM

**Testes de comparação múltipla post hoc.** Depois de determinar que existem diferenças entre as médias, testes de intervalo a posteriori e múltiplas comparações aos pares podem determinar quais médias diferem. As comparações são feitas em valores não ajustados. Esses testes são usados apenas para fatores fixos entre sujeitos. Em Medidas Repetidas GLM, esses testes não estão disponíveis se não houver fatores entre sujeitos, e os testes de múltipla comparação a posteriori são realizados para a média entre os níveis dos fatores entre sujeitos. Para GLM Multivariado, os testes a posteriori são realizados para cada variável dependente separadamente. GLM multivariado e Medidas Repetidas do GLM estão disponíveis apenas se você tiver a opção Advanced Statistics instalada.

Os testes de diferença honestamente significativa de Bonferroni e Tukey são comumente usados como testes de comparação múltipla. O **teste de Bonferroni**, baseado na estatística  $t$  do aluno, ajusta o nível de significância observado para o fato de que comparações múltiplas são feitas. O **teste  $t$  de Sidak** também ajusta o nível de significância e fornece limites mais apertados do que o teste de Bonferroni. O **teste da diferença significativa honesta de Tukey** usa o intervalo estatístico estudentizado para fazer todas as comparações entre pares entre os grupos e define a taxa de erro entre experimentos na taxa de erros da coleção de todas as comparações entre pares. Ao testar um grande número de pares de médias, o teste de diferença honestamente significativa de Tukey é mais poderoso que o teste de Bonferroni. Para um pequeno número de pares, Bonferroni é mais poderoso.

**GT2 de Hochberg** é semelhante ao teste da diferença significativa honesta de Tukey, mas o módulo máximo estudentizado é utilizado. Normalmente, o teste de Tukey é mais poderoso. O **teste de comparação entre pares de Gabriel** também utiliza o módulo máximo estudentizado e é geralmente

mais poderoso que o GT2 de Hochberg quando os tamanhos de células são desiguais. O teste de Gabriel pode se tornar liberal quando os tamanhos das células variam enormemente.

O **teste t de comparação múltipla entre pares de Dunnett** compara um conjunto de tratamentos com relação a uma média de controle única. A última categoria é a categoria de controle padrão. Como alternativa, é possível escolher a primeira categoria. Também é possível escolher um teste de dois lados ou unilateral. Para testar se a média em qualquer nível (exceto na categoria de controle) do fator não é igual à da categoria de controle, use um teste bilateral. Para testar se a média em qualquer nível do fator é menor do que a da categoria de controle, selecione **< Controle**. Da mesma forma, para testar se a média em qualquer nível do fator é maior do que a da categoria de controle, selecione **> Controle**.

Ryan, Einot, Gabriel e Welsch (R-E-G-W) desenvolveram dois testes de alcance múltiplo de redução. Múltiplos procedimentos de redução primeiro testam se todas as médias são iguais. Se todas as médias não forem iguais, os subconjuntos de médias são testados quanto à igualdade. **R-E-G-W F** é baseado em um teste *F*, e **R-E-G-W Q** é baseado no intervalo estudentizado. Esses testes são mais poderosos do que o teste de amplitude múltipla de Duncan e o Student-Newman-Keuls (que também são procedimentos de redução múltipla), mas não são recomendados para tamanhos de células desiguais.

Quando as variâncias são desiguais, utilize **T2 de Tamhane** (testes de comparações entre pares conservadores com base em um teste *t*), **T3 de Dunnett** (teste de comparação entre pares com base no módulo máximo estudentizado), **Games-Howell Teste de comparação entre pares** (às vezes liberal) ou **C de Dunnett** (teste de comparação entre pares com base no intervalo estudentizado). Observe que esses testes não são válidos e não serão produzidos se houver vários fatores no modelo.

**Teste de amplitude múltipla de Duncan**, Student-Newman-Keuls (**S-N-K**) e **b de Tukey** são testes que ranqueiam médias de grupo e calculam um valor de intervalo. Esses testes não são usados com tanta frequência quanto os testes discutidos anteriormente.

O **teste t de Waller-Duncan** usa uma abordagem bayesiana. Este teste de intervalo usa a média harmônica do tamanho da amostra quando os tamanhos das amostras são desiguais.

O nível de significância do teste **Scheffé** foi projetado para permitir que todas as combinações lineares possíveis das médias de grupo fossem testadas, não apenas as comparações entre pares disponíveis neste recurso. O resultado é que o teste de Scheffé é muitas vezes mais conservador do que outros testes, o que significa que uma diferença maior entre as médias é necessária para significância.

O teste de múltipla comparação entre pares da diferença menos significativa (**LSD**) é equivalente a testes *t* individuais múltiplos entre todos os pares de grupos. A desvantagem deste teste é que nenhuma tentativa é feita para ajustar o nível de significância observado para comparações múltiplas.

**Testes exibidos.** As comparações entre pares são fornecidas para LSD, Sidak, Bonferroni, Games-Howell, T2 e T3 de Tamhane, Cde Dunnett e T3 de Dunnett. Subconjuntos homogêneos para testes de intervalo são fornecidos para S-N-K, *b* de Tukey, Duncan, *F* de R-E-G-W, *Q* de R-E-G-W e Waller. O teste da diferença significativa honesta de Tukey, o GT2 de Hochberg, o teste de Gabriel e o teste de Scheffé são testes de comparação múltipla e testes de intervalo.

## Opções de GLM

Estatísticas de opcionais estão disponíveis a partir dessa caixa de diálogo. As estatísticas são calculadas utilizando um modelo de efeitos fixos.

**Exibição.** Selecione **Estatísticas descritivas** para produzir médias observadas, desvios padrão e contagens para todas as variáveis dependentes em todas as células. **Estimativas de tamanho de efeito** fornece um valor *eta* quadrado parcial para cada efeito e cada estimativa paramétrica. A estatística *eta* quadrado descreve a proporção da variabilidade total atribuível a um fator. Selecione **Potência observada** para obter a potência do teste quando a hipótese alternativa é configurada com base no valor observado. Selecione **Estimativas paramétrica** para produzir as estimativas paramétrica, os erros padrão, testes *t*, intervalos de confiança e a potência observada para cada teste. Selecione **Matriz de coeficientes de contraste** para obter a matriz **L**.

Os **Testes de homogeneidade** produzem testes de Levene da homogeneidade da variância para cada variável dependente em todas as combinações de nível dos fatores entre assuntos, somente para fatores entre assuntos. As opções de gráficos de dispersão versus nível e de resíduos são úteis para verificar

suposições sobre os dados. Esse item ficará desativado se não houver fatores. Selecione **Gráfico de resíduo** para produzir um gráfico residual observado por predito por padronizado para cada variável dependente. Esses gráficos são úteis para investigar a suposição de variância igual. Selecione **Falta de ajuste** para verificar se o relacionamento entre a variável dependente e as variáveis independentes pode ser descrito adequadamente pelo modelo. **Funções estimáveis gerais** permite construir testes de hipótese customizados com base nas funções estimáveis gerais. As linhas em qualquer matriz de coeficiente de contraste são combinações lineares das funções estimáveis gerais.

Os **Testes de heterocedasticidade** estão disponíveis para testar se a variância dos erros (para cada variável dependente) depende dos valores das variáveis independentes. Para o **Teste de Breusch-Pagan**, **Teste de Breusch-Pagan modificado** e o **Teste F**, é possível especificar o modelo no qual o teste é baseado. Por padrão, o modelo consiste em um termo constante, um termo que é linear nos valores preditos, um termo que é quadrático nos valores preditos e um termo de erro.

As **Estimativas de parâmetro com erros padrão robustos** exibem uma tabela de estimativas de parâmetro, juntamente com erros padrão robustos ou consistentes com heterocedasticidade (HC); além de estatísticas de  $t$ , valores de significância e intervalos de confiança que usam os erros padrão robustos. Há cinco métodos diferentes disponíveis para a estimativa da matriz de covariâncias robusta.

#### HC0

Baseado no assintótico original ou no estimador robusto de amostra grande, empírico ou "sanduíche" da matriz de covariâncias das estimativas de parâmetro. A parte do meio do sanduíche contém resíduos de OLS (quadrados mínimos ordinários) quadrados ou de WSL (quadrados mínimos ponderados) ponderados quadrados.

#### HC1

Uma modificação de amostra finita de HC0, multiplicada por  $N/(N-p)$ , em que  $N$  é o tamanho da amostra e  $p$  é o número de parâmetros não redundantes no modelo.

#### HC2

Uma modificação de HC0 que envolve a divisão do resíduo quadrado por  $1-h$ , em que  $h$  é a alavanca para o caso.

#### HC3

Uma modificação de HC0 que aproxima um estimador jackknife. Os resíduos quadrados são divididos pelo quadrado de  $1-h$ .

#### HC4

Uma modificação de HC0 que divide os resíduos quadrados por  $1-h$  a uma potência que varia de acordo com  $h$ ,  $N$  e  $p$ , com um limite superior de 4.

**Nível de Significância.** Talvez você queira ajustar o nível de significância utilizado em testes posteriori e o nível de confiança utilizado para construir intervalos de confiança. O valor especificado é também usado para calcular o poder observado para o teste. Ao especificar um nível de significância, o nível associado dos intervalos de confiança é exibido na caixa de diálogo.

## Recursos adicionais do comando UNIANOVA

O idioma da sintaxe de comando também permite:

- Especificar efeitos aninhados no design (utilizando o subcomando DESIGN).
- Especificar testes de efeitos versus uma combinação linear de efeitos ou um valor (utilizando o subcomando TEST).
- Especificar diversos contrastes (utilizando o subcomando CONTRAST).
- Incluir valores omissos de usuário (utilizando o subcomando MISSING).
- Especificar critérios do ESP (utilizando o subcomando CRITERIA).
- Construir uma matriz **L**, uma matriz **M** ou uma matriz **K** customizada (usando os subcomandos LMATRIX, MMATRIX e KMATRIX).
- Para desvio ou contrastes simples, especifique uma categoria de referência intermediária (utilizando o subcomando CONTRAST).
- Especificar métricas para contrastes polinomiais (utilizando o subcomando CONTRAST).

- Especificar termos de erro para comparações post hoc (utilizando o subcomando POSTHOC).
- Calcular médias marginais estimadas para qualquer fator ou interação entre fatores entre os fatores na lista de fatores (utilizando o subcomando EMMEANS).
- Especificar nomes para variáveis temporárias (utilizando o subcomando SAVE).
- Construir um arquivo de dados de matriz de correlações (utilizando o subcomando OUTFILE).
- Construir um arquivo de dados de matriz que contenha estatísticas da tabela ANOVA entre assuntos (utilizando o subcomando OUTFILE).
- Salvar a matriz de design em um novo arquivo de dados (utilizando o subcomando OUTFILE).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Salvamento de GLM

É possível salvar valores preditos pelo modelo, resíduos e medidas relacionadas às novas variáveis no Editor de dados. Muitas dessas variáveis podem ser usadas para examinar suposições a respeito dos dados. Para salvar os valores para usar em outra sessão do IBM SPSS Statistics, deve-se salvar o arquivo de dados atual

**Valores Preditos.** Os valores que o modelo prediz para cada caso.

- *Não padronizado.* O valor que o modelo prediz para a variável dependente.
- *Ponderado.* Valores Preditos Não Padronizados Ponderados Disponível apenas se uma variável WLS foi selecionada anteriormente.
- *Erro padrão.* Uma estimativa do desvio padrão do valor médio da variável dependente para os casos que tiverem os mesmos valores das variáveis independentes.

**Diagnósticos.** Medidas para identificar os casos com combinações de valores incomuns para as variáveis independentes e os casos que podem ter um grande impacto no modelo.

- *Distância de Cook.* Uma medida de quanto os resíduos de todos os casos mudariam se um caso em particular fosse excluído do cálculo dos coeficientes de regressão. Um D de Cook grande indica que excluir um caso do cálculo das estatísticas de regressão altera os coeficientes substancialmente.
- *Valores de ponto de alavanca.* Valores de alavancagem não centralizados A influência relativa de cada observação sobre o ajuste do modelo.

**Residuais.** Um resíduo não padronizado é o valor real da variável dependente menos o valor predito pelo modelo. Resíduos padronizados, estudentizados e excluídos também estão disponíveis. Se uma variável WLS foi escolhida, os resíduos não padronizados ponderados estão disponíveis.

- *Não padronizado.* A diferença entre um valor observado e o valor predito pelo modelo
- *Ponderado.* Resíduos não padronizados ponderados Disponível apenas se uma variável WLS foi selecionada anteriormente.
- *Padronizado.* O resíduo dividido por uma estimativa de seu desvio padrão Resíduos padronizados, também conhecidos como resíduos de Pearson, possuem uma média de 0 e um desvio padrão de 1.
- *Estudentizado.* O residual dividido por uma estimativa de seu desvio-padrão que varia de caso para caso, dependendo da distância de valores de cada caso sobre as variáveis independentes a partir dos meios das variáveis independentes. Às vezes referido como resíduos estudados internamente.
- *Excluído.* O resíduo para um caso em que esse caso é excluído do cálculo dos coeficientes de regressão Ele é a diferença entre o valor da variável dependente e o valor predito ajustado.

**Estatísticas de coeficiente.** Grava uma matriz de variância-covariância das estimativas do parâmetro no modelo em um novo conjunto de dados na sessão atual ou em um arquivo de dados externo do IBM SPSS Statistics. Além disso, para cada variável dependente, haverá uma linha de estimativas de parâmetro, uma linha de erros padrão das estimativas de parâmetro, uma linha de valores de significância para as estatísticas *t* correspondentes às estimativas de parâmetros e uma linha de graus de liberdade de resíduos. Para um modelo multivariado, há linhas semelhantes para cada variável dependente. Quando a estatística consistente com heteroscedasticidade é selecionada (disponível apenas para modelos univariados), a matriz de variância-covariância é calculada usando um estimador robusto, a linha de erros

padrão exibe os erros padrão robustos e os valores de significância refletem os erros robustos. É possível usar este arquivo de matriz em outros procedimentos que lêem arquivos de matriz.

## Médias marginais estimadas de GLM

Selecione os fatores e interações para os quais deseja estimativas das médias marginais da população nas células. Essas médias são ajustadas para as covariáveis, se houver.

### Comparar efeitos principais

Fornecer comparações pareadas não corrigidas entre as médias marginais estimadas para qualquer efeito principal no modelo, tanto para fatores entre e dentro dos sujeitos. Este item está disponível somente se os efeitos principais forem selecionados na lista Exibir meios para.

### Comparar efeitos principais simples

A configuração é ativada sempre que a lista de destino contém um ou mais efeitos de produto ou de interação (por exemplo,  $A*B$ ,  $A*B*C$ ). A configuração suporta a especificação de comparações entre efeitos principais simples, que são efeitos principais aninhados nos níveis de outros fatores.

### Ajuste do intervalo de confiança

Selecione a diferença menos significativa (LSD), Bonferroni ou ajuste de Sidak para os intervalos de confiança e significância. Este item fica disponível apenas ao selecionar **Comparar efeitos principais** e/ou **Comparar efeitos principais simples**.

## Especificando médias marginais estimadas

1. Nos menus, escolha um dos procedimentos disponíveis em > **Analisar** > **Modelo linear geral**.
2. No diálogo principal, clique em **Médias de EM**.

## Opções de GLM

Estatísticas de opcionais estão disponíveis a partir dessa caixa de diálogo. As estatísticas são calculadas utilizando um modelo de efeitos fixos.

**Exibição.** Selecione **Estatísticas descritivas** para produzir médias observadas, desvios padrão e contagens para todas as variáveis dependentes em todas as células. **Estimativas de tamanho de efeito** fornece um valor eta quadrado parcial para cada efeito e cada estimativa paramétrica. A estatística eta quadrado descreve a proporção da variabilidade total atribuível a um fator. Selecione **Potência observada** para obter a potência do teste quando a hipótese alternativa é configurada com base no valor observado. Selecione **Estimativas paramétrica** para produzir as estimativas paramétrica, os erros padrão, testes  $t$ , intervalos de confiança e a potência observada para cada teste. Selecione **Matriz de coeficientes de contraste** para obter a matriz **L**.

Os **Testes de homogeneidade** produzem testes de Levene da homogeneidade da variância para cada variável dependente em todas as combinações de nível dos fatores entre assuntos, somente para fatores entre assuntos. As opções de gráficos de dispersão versus nível e de resíduos são úteis para verificar suposições sobre os dados. Esse item ficará desativado se não houver fatores. Selecione **Gráfico de resíduo** para produzir um gráfico residual observado por predito por padronizado para cada variável dependente. Esses gráficos são úteis para investigar a suposição de variância igual. Selecione **Falta de ajuste** para verificar se o relacionamento entre a variável dependente e as variáveis independentes pode ser descrito adequadamente pelo modelo. **Funções estimáveis gerais** permite construir testes de hipótese customizados com base nas funções estimáveis gerais. As linhas em qualquer matriz de coeficiente de contraste são combinações lineares das funções estimáveis gerais.

Os **Testes de heterocedasticidade** estão disponíveis para testar se a variância dos erros (para cada variável dependente) depende dos valores das variáveis independentes. Para o **Teste de Breusch-Pagan**, **Teste de Breusch-Pagan modificado** e o **Teste F**, é possível especificar o modelo no qual o teste é baseado. Por padrão, o modelo consiste em um termo constante, um termo que é linear nos valores preditos, um termo que é quadrático nos valores preditos e um termo de erro.

As **Estimativas de parâmetro com erros padrão robustos** exibem uma tabela de estimativas de parâmetro, juntamente com erros padrão robustos ou consistentes com heterocedasticidade (HC); além

de estatísticas de  $t$ , valores de significância e intervalos de confiança que usam os erros padrão robustos. Há cinco métodos diferentes disponíveis para a estimativa da matriz de covariâncias robusta.

#### **HC0**

Baseado no assintótico original ou no estimador robusto de amostra grande, empírico ou "sanduíche" da matriz de covariâncias das estimativas de parâmetro. A parte do meio do sanduíche contém resíduos de OLS (quadrados mínimos ordinários) quadrados ou de WSL (quadrados mínimos ponderados) ponderados quadrados.

#### **HC1**

Uma modificação de amostra finita de HC0, multiplicada por  $N/(N-p)$ , em que  $N$  é o tamanho da amostra e  $p$  é o número de parâmetros não redundantes no modelo.

#### **HC2**

Uma modificação de HC0 que envolve a divisão do resíduo quadrado por  $1-h$ , em que  $h$  é a alavanca para o caso.

#### **HC3**

Uma modificação de HC0 que aproxima um estimador jackknife. Os resíduos quadrados são divididos pelo quadrado de  $1-h$ .

#### **HC4**

Uma modificação de HC0 que divide os resíduos quadrados por  $1-h$  a uma potência que varia de acordo com  $h$ ,  $N$  e  $p$ , com um limite superior de 4.

**Nível de Significância.** Talvez você queira ajustar o nível de significância utilizado em testes posteriori e o nível de confiança utilizado para construir intervalos de confiança. O valor especificado é também usado para calcular o poder observado para o teste. Ao especificar um nível de significância, o nível associado dos intervalos de confiança é exibido na caixa de diálogo.

## **Modelo de regressão auxiliar GLM**

A caixa de diálogo Modelo de regressão auxiliar especifica o modelo usado para testar a heterocedasticidade.

#### **Usar valores preditos**

Usa um modelo que consiste em um termo constante, um termo linear nos valores previstos, um termo quadrático nos valores previstos e um termo de erro.

#### **Usar modelo univariado**

Utiliza o modelo especificado no subdiálogo Modelo. Um termo de interceptação é incluído se o modelo especificado não contiver um.

#### **Modelo customizado**

Utiliza o modelo especificado explicitamente.

#### **Termos de construção**

Use essa opção quando desejar incluir termos não aninhados de um determinado tipo (como efeitos principais) para todas as combinações de um conjunto selecionado de fatores e covariáveis.

#### **Construir termos customizados**

Use esta opção quando desejar incluir termos aninhados ou quando desejar construir explicitamente qualquer variável de termo por variável. A criação de um termo aninhado envolve as etapas a seguir:

## **Recursos adicionais do comando UNIANOVA**

O idioma da sintaxe de comando também permite:

- Especificar efeitos aninhados no design (utilizando o subcomando DESIGN).
- Especificar testes de efeitos versus uma combinação linear de efeitos ou um valor (utilizando o subcomando TEST).
- Especificar diversos contrastes (utilizando o subcomando CONTRAST).

- Incluir valores omissos de usuário (utilizando o subcomando MISSING).
- Especificar critérios do ESP (utilizando o subcomando CRITERIA).
- Construir uma matriz **L**, uma matriz **M** ou uma matriz **K** customizada (usando os subcomandos LMATRIX, MMATRIX e KMATRIX).
- Para desvio ou contrastes simples, especifique uma categoria de referência intermediária (utilizando o subcomando CONTRAST).
- Especificar métricas para contrastes polinomiais (utilizando o subcomando CONTRAST).
- Especificar termos de erro para comparações post hoc (utilizando o subcomando POSTHOC).
- Calcular médias marginais estimadas para qualquer fator ou interação entre fatores entre os fatores na lista de fatores (utilizando o subcomando EMMEANS).
- Especificar nomes para variáveis temporárias (utilizando o subcomando SAVE).
- Construir um arquivo de dados de matriz de correlações (utilizando o subcomando OUTFILE).
- Construir um arquivo de dados de matriz que contenha estatísticas da tabela ANOVA entre assuntos (utilizando o subcomando OUTFILE).
- Salvar a matriz de design em um novo arquivo de dados (utilizando o subcomando OUTFILE).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Correlações bivariadas

O procedimento Correlações Bivariadas calcula o coeficiente de correlação de Pearson, o  $r$  de Spearman e o  $\tau$ - $b$  de Kendall com seus níveis de significância. As correlações medem como as ordens de variáveis ou de ranqueamento estão relacionadas. Antes de calcular um coeficiente de correlação, analise seus dados em busca de valores discrepantes (que podem causar resultados enganosos) e evidências de uma relação linear. O coeficiente de correlação de Pearson é uma medida de associação linear. Duas variáveis podem estar perfeitamente relacionadas, mas se a relação não for linear, o coeficiente de correlação de Pearson não é uma estatística apropriada para medir sua associação.

As configurações de intervalo de confiança estão disponíveis para Pearson e Spearman.

### Exemplo

O número de jogos ganhos por um time de basquete está correlacionado com o número médio de pontos marcados por jogo? Um gráfico de dispersão indica que há um relacionamento linear. Analisando os dados da temporada 1994-1995 da NBA, o coeficiente de correlação de Pearson (0,581) é significativo no nível de 0,01. Você pode suspeitar que quanto mais jogos ganhos por temporada, menos pontos os oponentes marcaram. Essas variáveis são correlacionadas negativamente (-0,401), e a correlação é significativa no nível de 0,05.

### Estatísticas

Para cada variável: número de casos com valores não omissos, média e desvio padrão. Para cada par de variáveis: coeficiente de correlação de Pearson,  $r$  de Spearman,  $\tau$ - $b$  de Kendall, produto cruzado de desvios e covariâncias.

## Considerações de dados

### Dados

Utilize variáveis quantitativas simétricas para o coeficiente de correlação de Pearson e variáveis quantitativas ou variáveis com categorias ordenadas por  $r$  de Spearman e  $\tau$ - $b$  de Kendall.

### Suposições

O coeficiente de correlação de Pearson assume que cada par de variáveis é normal bivariado.

## Obtendo correlações bivariadas

A partir dos menus, escolha:

**Analisar > Correlacionar > Bivariado...**

1. Selecione duas ou mais variáveis numéricas.

As opções a seguir também estão disponíveis:

#### **Coefficientes de correlação**

Para variáveis quantitativas, distribuídas normalmente, escolha o coeficiente de correlação **Pearson**. Se seus dados não estiverem normalmente distribuídos ou possuírem categorias ordenadas, escolha **tau-b de Kendall** ou **Spearman**, que mede a associação entre as ordens de ranqueamento. Os coeficientes de correlação variam em valor de  $-1$  (uma relação negativa perfeita) e  $+1$  (uma relação positiva perfeita). Um valor de  $0$  indica nenhum relacionamento linear. Ao interpretar seus resultados, tome cuidado para não tirar conclusões de causa e efeito devido a uma correlação significativa.

#### **Teste de significância**

É possível selecionar probabilidades bicaudais ou unicaudais. Se a direção da associação for conhecida com antecedência, selecione **Unilateral**. Caso contrário, selecione **Bilateral**.

#### **Sinalizar correlações significativas**

Os coeficientes de correlação significativos no nível  $0,05$  são identificados com um único asterisco, e os significativos no nível  $0,01$  são identificados com dois asteriscos.

#### **Mostrar somente o triângulo inferior**

Quando selecionado, apenas o triângulo inferior da tabela de matriz de correlação é apresentado na saída. Quando não selecionado, a tabela de matriz de correlação completa é apresentada na saída. A configuração permite que a saída de tabela siga as diretrizes de estilo de APA.

#### **Mostrar diagonal**

Quando selecionado, o triângulo inferior da tabela de matriz de correlação junto com os valores diagonais são apresentados na saída. A configuração permite que a saída de tabela siga as diretrizes de estilo de APA.

2. É possível selecionar opcionalmente o seguinte:

- Clique em **Opções...** para especificar estatísticas de correlação de Pearson e configurações de valores omissos.
- Clique em **Estilo...** para especificar condições para mudar automaticamente as propriedades das tabelas dinâmicas com base em condições específicas.
- Clique em **Autoinicialização...** para derivar estimativas robustas de erros padrão e intervalos de confiança para estimativas como a média, mediana, proporção, razão de chances, coeficiente de correlação ou coeficiente de regressão.
- Clique em **Intervalo de confiança...** para definir as opções para a estimação dos intervalos de confiança.

## **Opções de correlações bivariadas**

### **Estatísticas**

Para as correlações Pearson, é possível escolher um ou ambos os seguintes:

#### **Médias e desvios padrão**

Exibida para cada variável. O número de casos com valores não omissos também é mostrado. Os valores ausentes são tratados variável por variável, independentemente de sua configuração de valores ausentes.

#### **Desvios de produto vetorial e covariâncias**

Exibido para cada par de variáveis. O produto cruzado dos desvios é igual à soma dos produtos das variáveis corrigidas pela média. Este é o numerador do coeficiente de correlação de Pearson. A covariância é uma medida não padronizada do relacionamento entre duas variáveis, igual ao desvio de produto vetorial dividido por  $N-1$ .

### **Valores omissos**

É possível escolher um dos seguintes:



### Excluir casos de pairwise

Casos com valores ausentes para uma ou ambas de um par de variáveis para um coeficiente de correlação são excluídos da análise. Como cada coeficiente é baseado em todos os casos que possuem códigos válidos naquele par de variáveis em particular, a informação máxima disponível é usada em cada cálculo. Isso pode resultar em um conjunto de coeficientes com base em um número variável de casos.

### Excluir casos de listwise

Casos com valores ausentes para qualquer variável são excluídos de todas as correlações.

## Intervalo de confiança de bivariate correlations

O diálogo Intervalo de Confiança fornece opções para a estimação dos intervalos de confiança. O diálogo está disponível quando **Pearson**, **Kendall's tau-b** ou **Spearman** é selecionado no diálogo Bivariate correlation.

### Estimar intervalo de confiança do parâmetro bivariate correlation

Controla a estimação de intervalo de confiança do parâmetro de correlação bivariada. Quando selecionada, a estimação do intervalo de confiança ocorre.

### Intervalo de confiança (%)

Especifica o nível de confiança para todos os intervalos de confiança produzidos. Especifique um valor numérico entre 0 e 100. O valor padrão é 95.

### Correlação de Pearson

A configuração **Aplicar o ajuste de viés** controla se o ajuste de viés é aplicado. Por padrão, a configuração não é selecionada, o que não leva em consideração o termo de viés. Quando selecionado, o ajuste de viés para a estimação dos limites de confiança é aplicado. A configuração está disponível quando **Pearson** é selecionado no diálogo Bivariate correlation.

### Correlação Spearman

A configuração está disponível quando **Spearman** é selecionado no diálogo Bivariate correlation e fornece opções para estimar a variância da Correlação Spearman por meio dos métodos a seguir:

- **Fieller, Hartley e Pearson**
- **Bonett e Wright**
- **Coruso e Cliff**

## Recursos adicionais dos comandos CORRELATIONS e NONPAR CORR

O idioma da sintaxe de comando também permite:

- Gravar uma matriz de correlações para correlações de Pearson que podem ser utilizadas no lugar de dados brutos para obter outras análises como análise fatorial (com o subcomando MATRIX).
- Obter correlações de cada variável em uma lista com cada variável em uma segunda lista (utilizando a palavra-chave WITH no subcomando VARIABLES).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Correlações parciais

O procedimento Correlações Parciais calcula coeficientes de correlação parcial que descrevem o relacionamento linear entre duas variáveis enquanto controla os efeitos de uma ou mais variáveis adicionais. As correlações são medidas de associação linear. Duas variáveis podem estar perfeitamente relacionadas, mas se a relação não for linear, um coeficiente de correlação não é uma estatística apropriada para medir sua associação.

### Exemplo

Existe uma relação entre financiamento de assistência médica e taxas de doença? Embora talvez você espere que qualquer relacionamento desse tipo seja um relacionamento negativo, um estudo revela uma correlação *positiva* significativa: conforme o financiamento em assistência médica aumenta, taxas de doença parecem aumentar. Controlar a taxa de visitas aos profissionais de saúde, no

entanto, praticamente elimina a correlação positiva observada. O financiamento dos cuidados de saúde e as taxas de doença só parecem estar positivamente relacionados porque mais pessoas têm acesso aos cuidados de saúde quando o financiamento aumenta, o que leva a mais doenças relatadas por médicos e hospitais.

### **Estatísticas**

Para cada variável: número de casos com valores não omissos, média e desvio padrão. Matrizes de correlação parcial e de ordem zero, com graus de liberdade e níveis de significância.

## **Considerações de dados**

### **Dados**

Use variáveis quantitativas simétricas.

### **Suposições**

O procedimento de Correlações Parciais assume que cada par de variáveis é normal bivariado.

## **Obtendo correlações parciais**

1. A partir dos menus, escolha:

**Analisar > Correlacionar > Parcial ...**

2. Selecione duas ou mais variáveis numéricas para as quais as correlações parciais devem ser calculadas.
3. Selecione uma ou mais variáveis de controle numéricas.

As opções a seguir também estão disponíveis:

### **Teste de significância**

É possível selecionar probabilidades bicaudais ou unicaudais. Se a direção da associação for conhecida com antecedência, selecione **Unilateral**. Caso contrário, selecione **Bilateral**.

### **Exibir nível de significância real**

Por padrão, a probabilidade e os graus de liberdade são mostrados para cada coeficiente de correlação. Se você desmarcar este item, os coeficientes significativos no nível 0,05 são identificados com um único asterisco, os coeficientes significativos no nível 0,01 são identificados com um asterisco duplo e os graus de liberdade são suprimidos. Essa configuração afeta matrizes de correlação parcial e de ordem zero.

## **Opções de correlações parciais**

**Estatísticas.** É possível escolher um ou ambos os seguintes itens:

- **Médias e desvios padrão.** Exibida para cada variável. O número de casos com valores não omissos também é mostrado.
- **Correlações de ordem zero.** Uma matriz de correlações simples entre todas as variáveis, incluindo variáveis de controle, é exibida.

**Valores omissos.** É possível escolher uma das alternativas a seguir:

- **Excluir listwise dos casos.** Casos com valores ausentes para qualquer variável, incluindo uma variável de controle, são excluídos de todos os cálculos.
- **Excluir casos entre pares.** Para o cálculo das correlações de ordem zero nas quais as correlações parciais se baseiam, não é usado um caso com valores omissos para ambas ou uma de um par de variáveis. A exclusão em pares usa o máximo de dados possível. No entanto, o número de casos pode diferir entre os coeficientes. Quando a exclusão de pares está em vigor, os graus de liberdade para um determinado coeficiente parcial são baseados no menor número de casos usados no cálculo de qualquer uma das correlações de ordem zero.

## **Recursos adicionais do comando PARTIAL CORR**

O idioma da sintaxe de comando também permite:

- Ler uma matriz de correlações de ordem zero ou gravar uma matriz de correlações parciais (com o subcomando MATRIX).
- Obter correlações parciais entre duas listas de variáveis (utilizando a palavra-chave WITH no subcomando VARIABLES).
- Obter análises múltiplas (com diversos subcomandos VARIABLES).
- Especificar valores do pedido para solicitação (por exemplo, correlações parciais de primeira e segunda ordem) quando houver duas variáveis de controle (com o subcomando VARIABLES).
- Suprimir coeficientes redundantes (com o subcomando FORMAT).
- Exibir uma matriz de correlações simples quando alguns coeficientes não podem ser calculados (com o subcomando STATISTICS).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Distâncias

Este procedimento calcula qualquer variedade ampla de estatísticas que medem similaridades ou dissimilaridades (distâncias), seja entre pares de variáveis ou entre pares de caso. Essas medidas de similaridade ou de distância podem então ser utilizadas com outros procedimentos, como análise fatorial, análise de cluster ou ajuste de escala multidimensional, para ajudar a analisar conjuntos de dados complexos.

**Exemplo.** É possível medir semelhanças entre pares de automóveis com base em certas características, como tamanho do motor, MPG e potência? Ao calcular semelhanças entre automóveis, é possível ter uma noção de quais automóveis são semelhantes entre si e quais são diferentes entre si. Para uma análise mais formal, é possível considerar aplicar uma análise de cluster hierárquica ou dimensionamento multidimensional às semelhanças para explorar a estrutura subjacente.

**Estatísticas.** As medidas de dissimilaridade (distância) para dados de intervalo são distância euclidiana, distância euclidiana quadrada, Chebychev, bloco, Minkowski ou personalizadas; para dados de contagem, qui-quadrado ou phi-quadrado; para dados binários, distância euclidiana, distância euclidiana quadrada, diferença de tamanho, diferença de padrão, variância, forma ou Lance e Williams. As medidas de similaridade para dados do intervalo são correlação ou cosseno de Pearson; para dados binários, Russel e Rao, correspondência simples, Jaccard, divisão, Rogers e Tanimoto, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Kulczynski 1, Kulczynski 2, Sokal e Sneath 4, Hamann, Lambda, *D* de Anderberg, *Y* Yule, *Q* de Yuli, Ochiai, Sokal e Sneath 5, correlação de fi de 4 pontos ou dispersão.

Para Obter Matrizes de Distância

1. A partir dos menus, escolha:

**Analisar > Correlacionar > Distâncias...**

2. Selecione pelo menos uma variável numérica para calcular as distâncias entre os casos ou selecione pelo menos duas variáveis numéricas para calcular as distâncias entre as variáveis.
3. Selecione uma alternativa no grupo Calcular Distâncias para calcular as proximidades entre casos ou entre variáveis.

## Medidas de dissimilaridade de distâncias

No grupo Medida, selecione a alternativa que corresponde ao seu tipo de dados (intervalo, contagem ou binário); em seguida, na lista suspensa, selecione uma das medidas que corresponde a esse tipo de dados. As medidas disponíveis, por tipo de dados, são:

- **Dados do intervalo.** Distância euclidiana, distância euclidiana quadrada, Chebychev, bloco, Minkowski ou personalizado.
- **Dados de contagem.** Medida de qui-quadrado ou medida de fi-quadrado.
- **Dados binários.** Distância euclidiana, distância euclidiana ao quadrado, diferença de tamanho, diferença de padrão, variância, forma ou Lance e Williams. (Insira os valores para Presente e Ausente para especificar quais dois valores são significativos; Distâncias ignorará todos os outros valores.)

O grupo Valores de Transformação permite padronizar valores de dados para quaisquer casos ou variáveis antes do cálculo de proximidades. Essas transformações não são aplicáveis aos dados binários. Os métodos de padronização disponíveis são escores  $z$ , intervalo -1 a 1, intervalo 0 a 1, magnitude máxima de 1, média de 1 ou desvio padrão de 1.

O grupo Transformar Medidas permite transformar os valores gerados pela medida de distância. Eles são aplicados após a medida de distância ter sido calculada. As opções disponíveis são valores absolutos, alterar o sinal e redimensionar para o intervalo de 0–1.

## Medidas de similaridade de distâncias

No grupo Medida, selecione a alternativa que corresponde ao seu tipo de dado (intervalo ou binário); em seguida, na lista suspensa, selecione uma das medidas que corresponde a esse tipo de dados. As medidas disponíveis, por tipo de dados, são:

- **Dados do intervalo.** Correlação de Pearson ou cosseno.
- **Dados binários.** Russell e Rao, correspondência simples, Jaccard, Dice, Rogers e Tanimoto, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Kulczynski 1, Kulczynski 2, Sokal e Sneath 4, Hamann, Lambda,  $D$  d Anderberg,  $Y$  de Yule,  $Q$  de Yule, Ochiai, Sokal e Sneath 5, correlação de  $f_i$  de 4 pontos ou dispersão. (Insira os valores para Presente e Ausente para especificar quais dois valores são significativos; Distâncias ignorará todos os outros valores.)

O grupo Transformar valores permite padronizar valores de dados para casos ou variáveis antes de calcular as proximidades. Essas transformações não são aplicáveis aos dados binários. Os métodos de padronização disponíveis são escores  $z$ , intervalo -1 a 1, intervalo 0 a 1, magnitude máxima de 1, média de 1 e desvio padrão de 1.

O grupo Transformar Medidas permite transformar os valores gerados pela medida de distância. Eles são aplicados após a medida de distância ter sido calculada. As opções disponíveis são valores absolutos, mudar o sinal e redimensionar para o intervalo de 0–1.

## Recursos adicionais do comando PROXIMITIES

O procedimento Distâncias utiliza a sintaxe de comando PROXIMITIES. O idioma da sintaxe de comando também permite:

- Especifique qualquer número inteiro como a potência para a medida de distância de Minkowski.
- Especifique quaisquer números inteiros como a potência e raiz para uma medida de distância customizada.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Modelos lineares

---

Os modelos lineares preveem uma variável resposta contínua com base em relacionamentos lineares entre o destino e um ou mais preditores.

Os modelos lineares são relativamente simples e fornecem uma fórmula matemática de fácil interpretação para pontuação. As propriedades desses modelos são bem compreendidas e normalmente podem ser construídas muito rapidamente em comparação com outros tipos de modelos (como redes neurais ou árvores de decisão) no mesmo conjunto de dados.

**Exemplo.** Uma companhia de seguros com recursos limitados para investigar sinistros de seguros residenciais deseja construir um modelo para estimar custos de sinistros. Ao implantar esse modelo nos centros de serviços, os representantes podem inserir informações de sinistros ao telefone com um cliente e obter imediatamente o custo "esperado" do sinistro com base em dados anteriores.

**Requisitos de campo.** Deve haver um alvo e pelo menos uma entrada. Por padrão, os campos com funções predefinidas de Ambos ou Nenhum não são usados. O alvo deve ser contínuo (escala). Não há nenhuma restrição de nível de medição em preditores (entradas); campos (flag do nominal e ordinal) categóricos são usados como fatores no modelo e campos contínuos são usados como covariáveis

**Nota:** Se um campo categórico tiver mais de 1.000 categorias, o procedimento não será executado e nenhum modelo será construído.

## Para obter um modelo linear

Essa variável requer a opção de Base de Estatísticas.

A partir dos menus, escolha:

**Analisar > Regressão > Modelos Lineares Automáticos...**

1. Certifique-se de que exista pelo menos uma resposta e uma entrada.
2. Clique em **Opções de Criação** para especificar configurações de construção e de modelo opcionais.
3. Clique em **Opções de Modelo** para salvar escores no conjunto de dados ativo e exportar o modelo para um arquivo externo.
4. Clique em **Executar** para executar o procedimento e criar os objetos Modelo.

## Objetivos

**Qual é o seu principal objetivo?** Selecione o objetivo apropriado.

- **Crie um modelo padrão.** O método cria um único modelo para prever o destino usando os preditores. De um modo geral, os modelos padrão são mais fáceis de interpretar e podem ser pontuados mais rápido que combinações de conjuntos de dados impulsionados, ensacados ou grandes.
- **Aprimorar a precisão do modelo (boosting).** Cria um modelo de combinação usando o boosting, que gera uma sequência de modelos para obter predições mais precisas. Os conjuntos podem demorar mais para construir e pontuar do que um modelo padrão.

Boosting produz uma sucessão de "modelos de componentes", cada um dos quais é construído em todo o conjunto de dados. Antes de construir cada modelo de componente sucessivo, os registros são ponderados com base nos resíduos do modelo de componente anterior. Casos com grandes resíduos recebem pesos de análise relativamente mais altos para que o próximo modelo de componente se concentre em prever bem esses registros. Juntos, esses modelos de componentes formam um modelo de combinação. O modelo de combinação pontua novos registros usando uma regra de combinação; as regras disponíveis dependem do nível de medição do destino.

- **Aprimorar estabilidade do modelo (bagging).** O método constrói um modelo de combinação usando bagging (bootstrap agregador), que gera vários modelos para obter predições mais confiáveis. Os conjuntos podem demorar mais para construir e pontuar do que um modelo padrão.

A agregação de bootstrap (bagging) produz réplicas do conjunto de dados de treinamento por amostragem com substituição do conjunto de dados original. Isso cria amostras bootstrap de tamanho igual ao conjunto de dados original. Em seguida, um "modelo de componente" é construído em cada réplica. Juntos, esses modelos de componentes formam um modelo de combinação. O modelo de combinação pontua novos registros usando uma regra de combinação; as regras disponíveis dependem do nível de medição do destino.

- **Crie um modelo para conjuntos de dados muito grandes (requer o servidor IBM SPSS Statistics).** O método constrói um modelo de combinação dividindo o conjunto de dados em blocos de dados separados. Escolha esta opção se seu conjunto de dados for muito grande para construir qualquer um dos modelos acima ou para construção de modelo incremental. Essa opção pode levar menos tempo para ser construída, mas pode levar mais para ser pontuada do que um modelo padrão. Essa opção requer conectividade do IBM SPSS Statistics Server

Consulte [“Combinações”](#) na página 143 para obter configurações relacionadas a boosting, bagging e conjuntos de dados muito grandes boosting e bagging.

## Básicos

**Preparar dados automaticamente.** Essa opção ativa o procedimento para transformar internamente a resposta e os preditores para maximizar o poder preditivo do modelo; todas as transformações são

salvas com o modelo e aplicadas aos novos dados para pontuação. As versões originais dos campos transformados serão excluídas do modelo. Por padrão, a preparação automática de dados a seguir é executada.

- **Tratamento de Data e Hora.** Cada preditor de data é transformado em um novo preditor contínuo contendo o tempo decorrido desde uma data de referência (1970-01-01). Cada preditor de tempo é transformado em um novo preditor contínuo contendo o tempo decorrido desde um horário de referência (00:00:00).
- **Ajustar nível de medição.** Preditores contínuos com menos de 5 valores distintos são reformulados como preditores ordinais. Preditores ordinais com mais de 10 valores distintos são reformulados como preditores contínuos.
- **Tratamento de valor discrepante.** Valores de preditores contínuos que estão além de um valor de corte (3 desvios padrão a partir da média) são configurados para o valor de corte.
- **Tratamento de valor omissivo.** Os valores ausentes de preditores nominais são substituídos pelo modo da partição de treinamento. Os valores ausentes de preditores ordinais são substituídos pela mediana da partição de treinamento. Os valores ausentes de preditores contínuos são substituídos pela média da partição de treinamento.
- **Mesclagem supervisionada.** Isso torna um modelo mais econômico reduzindo o número de campos a serem processados em associação com a resposta. Categorias similares são identificadas com base no relacionamento entre a entrada e o destino. As categorias que não são significativamente diferentes (ou seja, com valor  $p$  maior que 0,1) são mescladas. Se todas as categorias forem mescladas em uma, as versões original e derivada do campo serão excluídas do modelo porque não têm valor como preditor.

**Nível de confiança.** Este é o nível de confiança utilizado para calcular estimativas de intervalo dos coeficientes do modelo na visualização [Coeficientes](#). Especifique um valor maior que 0 e menor que 100. O padrão é 95.

## Seleção de Modelo

**Método de seleção de modelo.** Escolha um dos métodos de seleção de modelo (detalhes abaixo) ou **Inclua todos os preditores**, o que simplesmente insere todos os preditores disponíveis como termos modelo dos principais efeitos. Por padrão, o **Forward stepwise** é utilizado.

**Seleção de Forward Stepwise.** Isso começa sem efeitos no modelo e adiciona e remove efeitos um passo de cada vez até que não seja mais possível adicionar ou remover de acordo com os critérios de stepwise.

- **Critérios de entrada/remoção.** Esta é a estatística usada para determinar se um efeito deve ser adicionado ou removido do modelo. O **Critério de Informações (AICC)** baseia-se na probabilidade do conjunto de treinamento dado o modelo, e é ajustado para penalizar modelos excessivamente complexos. As **Estatísticas de F** baseiam-se em um teste estatístico da melhoria no erro do modelo. **R-quadrado ajustado** é baseado no ajuste do conjunto de treinamento e é ajustado para penalizar modelos excessivamente complexos. **Critério de Prevenção ao Super Ajuste (ASE)** é baseado no ajuste (erro quadrático de média ou ASE) do conjunto de prevenção ao super ajuste. O conjunto de prevenção ao super ajuste é uma subamostra aleatória de aproximadamente 30% do conjunto de dados original que não é utilizado para treinar o modelo.

Se qualquer critério diferente de **Estatísticas F** for escolhido, então, em cada passo o efeito que corresponde ao maior aumento positivo no critério será incluído no modelo. Quaisquer efeitos no modelo que correspondam a uma diminuição no critério são removidos.

Se **Estatísticas de F** forem escolhidas como o critério, então, em cada passo, o efeito que tiver um valor de  $p$  menor que o limite especificado em **Incluir efeitos com valores de  $p$  menores que** será incluído no modelo. O padrão é 0,05. Quaisquer efeitos no modelo com um valor de  $p$  maior que o limite especificado em **Remover efeitos com valores de  $p$  maiores que** serão removidos. O padrão é 0,10.

- **Customizar o número máximo de efeitos no modelo final.** Por padrão, todos os efeitos disponíveis podem ser inseridos no modelo. Como alternativa, se o algoritmo stepwise terminar uma etapa com o número máximo de efeitos especificado, o algoritmo parará com o conjunto atual de efeitos.

- **Customizar número máximo de passos.** O algoritmo stepwise para após um determinado número de passos. Por padrão, isso é 3 vezes o número de efeitos disponíveis. Como alternativa, especifique um número inteiro máximo positivo de etapas.

**Seleção dos Melhores Subconjuntos.** Isso verifica "todos os modelos possíveis", ou pelo menos um subconjunto maior dos modelos possíveis do que forward stepwise, para escolher o melhor de acordo com o critério dos melhores subconjuntos. **O Critério de Informação (AICC)** é baseado na verossimilhança do conjunto de treinamento dado ao modelo e é ajustado para penalizar modelos excessivamente complexos. **R-quadrado ajustado** é baseado no ajuste do conjunto de treinamento e é ajustado para penalizar modelos excessivamente complexos. **Critério de Prevenção ao Super Ajuste (ASE)** é baseado no ajuste (erro quadrático de média ou ASE) do conjunto de prevenção ao super ajuste. O conjunto de prevenção ao super ajuste é uma subamostra aleatória de aproximadamente 30% do conjunto de dados original que não é utilizado para treinar o modelo.

O modelo com o maior valor do critério é escolhido como o melhor modelo.

**Nota:** A seleção dos melhores subconjuntos é computacionalmente mais intensiva do que a seleção forward stepwise. Quando os melhores subconjuntos são executados em conjunto com boosting, bagging ou conjuntos de dados muito grandes, pode levar muito mais tempo para construir do que um modelo padrão construído usando a seleção stepwise forward.

## Combinações

Essas configurações determinam o comportamento da combinação que ocorre quando boosting, bagging ou conjuntos de dados muito grandes são solicitados em Objetivos. Opções que não se aplicam ao objetivo selecionado são ignoradas.

**Bagging e Conjuntos de Dados Muito Grandes.** Ao escorar uma combinação, essa é a regra utilizada para combinar os valores preditos a partir dos modelos base para calcular o valor de escore de combinação.

- **Regra de combinação padrão para variáveis resposta contínuas.** Os valores preditos de combinação para variáveis resposta contínuas podem ser combinados utilizando a média ou mediana dos valores preditos a partir dos modelos base.

Observe que quando o objetivo é aprimorar a precisão do modelo, as seleções de regra de combinação são ignoradas. O boosting sempre usa uma votação majoritária ponderada para pontuar metas categóricas e uma mediana ponderada para pontuar metas contínuas.

**Boosting e Bagging.** Especifique o número de modelos base para construção quando o objetivo for melhorar a precisão ou a estabilidade do modelo; para bagging, este é o número de amostras de bootstrap. Ele deve ser um número inteiro positivo.

## Avançado

**Replicar resultados.** Configurar uma semente aleatória permite replicar análises. O gerador de número aleatório é utilizado para escolher quais registros estão no conjunto de prevenção ao super ajuste. Especifique um número inteiro ou clique em **Gerar**, que criará um pseudonúmero inteiro aleatório entre 1 e 2147483647, inclusive. O padrão é 54752075.

## Opções de Modelo

**Salvar valores preditos no conjunto de dados.** O nome de variável padrão é *PredictedValue*.

**Modelo de exportação.** Isso grava o modelo em um arquivo .zip externo. É possível usar esse arquivo de modelo para aplicar as informações de modelo a outros arquivos de dados para propósitos de escoragem. Especifique um nome de arquivo exclusivo e válido. Se a especificação do arquivo se referir a um arquivo existente, o arquivo será substituído.

## Sumarização do Modelo

A visualização de Resumo do Modelo é uma captura instantânea, um resumo rápido do modelo e seu ajuste.



**Tabela.** A tabela identifica algumas configurações de modelo de alto nível, incluindo:

- O nome da resposta especificado na guia Campos,
- Se a preparação de dado automático foi executada conforme especificado nas configurações Básicas,
- O método e o critério de seleção de modelo especificados nas configurações de Seleção de Modelo. O valor do critério de seleção para o modelo final também é exibido, e é apresentado em formato menor e melhor.

**Gráfico.** O gráfico exibe a precisão do modelo final, que é apresentada em formato maior, que é melhor. O valor é  $100 \times$  o valor de  $R^2$  ajustado para o modelo final.

## Preparação de Dado Automático

Essa exibição mostra informações sobre quais campos foram excluídos e como os campos transformados foram derivados na etapa de preparação automática de dados (ADP). Para cada campo que foi transformado ou excluído, a tabela lista o nome do campo, sua função na análise e a ação realizada pela etapa ADP. Os campos são classificados por ordem alfabética crescente dos nomes dos campos. As possíveis ações tomadas para cada campo incluem:

- **Duração da derivação: meses** calcula o tempo decorrido em meses a partir dos valores em um campo que contém datas para a data do sistema atual.
- **Duração da derivação: horas** calcula o tempo decorrido em horas a partir dos valores em um campo que contém tempos até o tempo atual do sistema.
- **Mudar nível de medição de contínuo para ordinal** reformula campos contínuos com menos de 5 valores exclusivos como campos ordinais.
- **Mudar nível de medição de ordinal para contínuo** reformula campos ordinais com mais de 10 valores exclusivos como campos contínuos.
- **Aparar valores discrepantes** configura valores de preditores contínuos que estão além de um valor de corte (3 desvios padrão a partir da média) para o valor de corte.
- **Substituir valores omissos** substitui valores omissos de campos nominais com os campos de modo, ordinais pelos campos de mediana e contínuos com a média.
- **Mesclar categorias para maximizar a associação com o destino** identifica categorias do preditor "semelhantes" com base no relacionamento entre a entrada e o destino. As categorias que não forem significativamente diferentes (ou seja, que possuem um valor de  $p$  maior que 0,05) são mescladas.
- **Excluir preditor de constante / após o tratamento de valor discrepante / após a combinação de categorias** remove preditores que têm um valor único, possivelmente após outras ações de ADP terem sido tomadas.

## Importância do Preditor

Normalmente, você desejará focar os seus esforços de modelagem nos campos preditores que importam mais e considerar descartar ou ignorar aqueles que menos importam. O gráfico de importância do preditor ajuda você a executar isso, indicando a importância relativa de cada preditor para estimar o modelo. Como os valores são relativos, a soma dos valores para todos os preditores na exibição é 1,0. A importância do preditor não é relacionada com a precisão do modelo. Ela é relacionada apenas com a importância de cada preditor em fazer uma previsão, não se a previsão é exata ou não.

## Predito por Observado

Isso exibe um gráfico de dispersão categorizado dos valores previstos no eixo vertical pelos valores observados no eixo horizontal. Idealmente, os pontos devem estar em uma linha de 45 graus; essa visualização pode lhe dizer se algum registro foi particularmente predito de forma inválida pelo modelo.

## Resíduos

Isso exibe um diagrama de diagnóstico dos resíduos de modelo.



**Estilos de gráfico.** Existem estilos de exibição diferentes, que são acessíveis a partir da lista suspensa de **Estilos**.

- **Histograma.** Esse é um histograma categorizado dos resíduos estudentizados com uma sobreposição da distribuição normal. Como os modelos lineares supõem que os resíduos têm uma distribuição normal, o histograma deverá se aproximar idealmente da linha suave.
- **Gráfico P-P.** Esse é um gráfico categorizado de probabilidade-probabilidade que compara os resíduos estudentizados com uma distribuição normal. Se a inclinação dos pontos plotados for menos acentuada do que a linha normal, os resíduos mostrarão uma variabilidade maior do que uma distribuição normal e, se a inclinação for mais acentuada, os resíduos mostrarão uma variabilidade menor do que uma distribuição normal. Se os pontos plotados tiverem uma curva em forma de S, a distribuição dos resíduos será defasada.

## Valores discrepantes

Esta tabela lista os registros que exercem influência indevida sobre o modelo e exibe o ID do registro (se especificado na guia Campos), o valor alvo e a distância de cook. A distância de cook é uma medida de quanto os resíduos de todos os registros mudariam se um determinado registro fosse excluído do cálculo dos coeficientes do modelo. Uma grande distância de cook indica que excluir um registro altera substancialmente os coeficientes e, portanto, deve ser considerado influente.

Os registros influentes devem ser examinados cuidadosamente para determinar se você pode dar a eles menos peso na estimativa do modelo, ou truncar os valores periféricos para algum limite aceitável ou remover completamente os registros influentes.

## Efeitos

Essa visualização exibe o tamanho de cada efeito no modelo.

**Estilos.** Existem estilos de exibição diferentes, que são acessíveis a partir da lista suspensa de **Estilos**.

- **Diagrama.** Esse é um gráfico no qual os efeitos são ordenados da parte superior até a inferior diminuindo a importância do preditor. As linhas de conexão no diagrama são ponderadas com base na significância do efeito, com a largura de linha maior correspondendo aos efeitos mais significativos (valores de  $p$  menores). Passar o mouse sobre a linha de conexão revela uma dica de ferramenta que mostra o valor de  $p$  e a importância do efeito. Este é o padrão.
- **Tabela.** Essa é uma tabela ANOVA para o modelo geral e os efeitos individuais do modelo. Os efeitos individuais são classificados de cima para baixo, diminuindo a importância do preditor. Observe que, por padrão, a tabela é reduzida para mostrar apenas os resultados do modelo geral. Para ver os resultados dos efeitos do modelo individual, clique na célula **Modelo Corrigido** na tabela.

**Importância do preditor.** Há uma régua de controle de Importância do Preditor que controla quais preditores são mostrados na visualização. Isso não altera o modelo, mas simplesmente permite que você se concentre nos preditores mais importantes. Por padrão, os 10 principais efeitos são exibidos.

**Significância.** Há uma régua de controle Significância que controla ainda mais quais efeitos são mostrados na visualização, além daqueles mostrados com base na importância do preditor. Efeitos com valores de significância maiores que o valor da régua de controle são ocultos. Isso não muda o modelo, mas simplesmente permite que você se concentre nos efeitos mais importantes. Por padrão o valor é 1,00, de modo que nenhum efeito seja filtrado com base na significância.

## Coefficientes

Essa visualização exibe o valor de cada coeficiente no modelo. Observe que fatores (preditores categóricos) são codificados pelo indicador dentro do modelo, de forma que **efeitos** contendo fatores geralmente terão múltiplos **coeficientes** associados; um para cada categoria, exceto a categoria correspondente ao parâmetro (referência) redundante.

**Estilos.** Existem estilos de exibição diferentes, que são acessíveis a partir da lista suspensa de **Estilos**.

- **Diagrama.** Esse é um gráfico que exibe o intercepto primeiro e, em seguida, ordena efeitos da parte superior até a inferior diminuindo a importância do preditor. Dentro de efeitos que contêm fatores, coeficientes são classificados por ordem crescente de valores de dados. As linhas de conexão no diagrama são coloridas com base no sinal do coeficiente (consulte a chave de diagrama) e ponderadas com base na significância do coeficiente, com a largura de linha maior correspondendo aos coeficientes mais significativos (valores de  $p$  menores). Passar o mouse sobre uma linha de conexão revela uma dica de ferramenta que mostra o valor do coeficiente, o seu valor  $p$  e a importância do efeito com o qual o parâmetro está associado. Esse é o estilo padrão.
- **Tabela.** Isso mostra os valores, testes de significância e intervalos de confiança para os coeficientes do modelo individual. Após a interceptação, os efeitos são classificados de cima para baixo, diminuindo a importância do preditor. Dentro de efeitos que contêm fatores, coeficientes são classificados por ordem crescente de valores de dados. Observe que, por padrão, a tabela é reduzida para mostrar apenas o coeficiente, a significância e a importância de cada parâmetro do modelo. Para ver o erro padrão, a estatística  $t$  e o intervalo de confiança, clique na célula de **Coeficiente** na tabela. Passar o mouse sobre o nome de um parâmetro de modelo na tabela revela uma dica de ferramenta que mostra o nome do parâmetro, o efeito ao qual o parâmetro está associado e (para preditores categóricos), os rótulos de valor associados ao parâmetro de modelo. Isso pode ser particularmente útil para ver as novas categorias criadas quando a preparação automática de dados mescla categorias semelhantes de um preditor categórico.

**Importância do preditor.** Há uma régua de controle de Importância do Preditor que controla quais preditores são mostrados na visualização. Isso não muda o modelo, mas simplesmente permite que você se concentre nos preditores mais importantes. Por padrão, os 10 principais efeitos são exibidos.

**Significância.** Há uma régua de controle de Significância que controla adicionalmente quais coeficientes são mostrados na visualização, além daqueles mostrados com base na importância do preditor. Coeficientes com valores de significância maiores que o valor da régua de controle são ocultados. Isso não altera o modelo, mas simplesmente permite que você se concentre nos coeficientes mais importantes. Por padrão o valor é 1,00, de modo que nenhum coeficiente seja filtrado com base na significância.

## Médias Estimadas

Estes são gráficos exibidos para preditores significativos. O gráfico exibe o valor estimado do modelo do destino no eixo vertical para cada valor do preditor no eixo horizontal, mantendo todos os outros preditores constantes. Ele fornece uma visualização útil dos efeitos dos coeficientes de cada preditor no destino.

*Nota:* se nenhum preditor for significativo, nenhuma média estimada será produzida.

## Sumarização de Construção de Modelo

Quando um algoritmo de seleção de modelo diferente de **Nenhum** é escolhido nas configurações de Seleção de Modelo, isso fornece alguns detalhes do processo de construção do modelo.

**Forward stepwise.** Quando forward stepwise é o algoritmo de seleção, a tabela exibe os últimos 10 passos no algoritmo stepwise. Para cada etapa, são mostrados o valor do critério de seleção e os efeitos no modelo naquela etapa. Isso lhe dá uma noção de quanto cada etapa contribui para o modelo. Cada coluna permite classificar as linhas para que você possa ver mais facilmente quais efeitos estão no modelo em uma determinada etapa.

**Melhores subconjuntos.** Quando melhores subconjuntos é o algoritmo de seleção, a tabela exibe os 10 principais modelos. Para cada modelo, são mostrados o valor do critério de seleção e os efeitos no modelo. Isto dá-lhe uma noção da estabilidade dos modelos de topo; se eles tendem a ter muitos efeitos semelhantes com algumas diferenças, então você pode estar bastante confiante no modelo "top"; se eles tendem a ter efeitos muito diferentes, alguns dos efeitos podem ser muito semelhantes e devem ser combinados (ou um deles removido). Cada coluna permite classificar as linhas para que você possa ver mais facilmente quais efeitos estão no modelo em uma determinada etapa.

## Regressão linear

---

A Regressão Linear estima os coeficientes da equação linear, envolvendo uma ou mais variáveis independentes, que melhor preveem o valor da variável dependente. Por exemplo, é possível tentar prever o total de vendas anual de um vendedor (a variável dependente) a partir de variáveis independentes como idade, educação e anos de experiência.

**Exemplo.** O número de jogos vencidos por um time de basquete em uma temporada está relacionado ao número médio de pontos que a equipe marca por jogo? Um gráfico de dispersão indica que essas variáveis estão linearmente relacionadas. O número de jogos vencidos e o número médio de pontos marcados pelo adversário também estão linearmente relacionados. Essas variáveis possuem um relacionamento negativo. Conforme o número de jogos vencidos aumenta, o número médio de pontos marcados pelo adversário diminui. Com a regressão linear, é possível modelar o relacionamento dessas variáveis. Um bom modelo pode ser utilizado para prever quantos jogos os times irão vencer.

**Estatísticas.** Para cada variável: número de casos válidos, média e desvio padrão. Para cada modelo: coeficientes de regressão, matriz de correlações, correlações de parte ou parciais,  $R$  múltiplo,  $R^2$ ,  $R^2$  ajustado, mudança em  $R^2$ , erro padrão da estimativa, tabela de análise de variância, valores preditos e resíduos. Além disso, intervalos de confiança de 95% para cada coeficiente de regressão, matriz de variância-covariância, fator de inflação de variância, tolerância, teste de Durbin-Watson, medidas de distância (de Mahalanobis, Cook e valores de ponto de alavanca), DfBeta, DfFit, intervalos de predição e informações de diagnóstico entre casos. Gráficos: de dispersão, gráficos parciais, histogramas e gráficos de probabilidade normal.

Considerações de Dados de Regressão Linear

**Dados.** As variáveis dependentes e independentes devem ser quantitativas. Variáveis categóricas, como religião, campo de estudo principal ou região de residência, precisam ser recodificadas para variáveis binárias (simuladas) ou para outros tipos de variáveis de contraste.

**Suposições.** Para cada valor da variável independente, a distribuição da variável dependente deve ser normal. A variância da distribuição da variável dependente deve ser constante para todos os valores da variável independente. O relacionamento entre a variável dependente e cada variável independente deve ser linear e todas as observações devem ser independentes.

Para Obter uma análise de regressão linear

1. A partir dos menus, escolha:

**Analisar > Regressão > Linear...**

2. Na caixa de diálogo Regressão Linear, selecione uma variável dependente numérica.

3. Selecione uma ou mais variáveis independentes numéricas.

Opcionalmente, é possível:

- Agrupar variáveis independentes em blocos e especificar métodos de entrada diferentes para diferentes subconjuntos de variáveis.
- Escolha uma variável de seleção para limitar a análise a um subconjunto de casos que possuem um valor específico para essa variável.
- Selecione uma variável de identificação de caso para identificar pontos em gráficos.
- Selecione uma variável numérica WLS Weight para uma análise de mínimos quadrados ponderados.

*WLS.* Permite obter um modelo de mínimos quadrados ponderados. Os pontos de dados são ponderados pelo recíproco de suas variâncias. Isso significa que as observações com variâncias grandes causam menos impacto na análise do que as observações associadas a variâncias pequenas. Se o valor da variável de ponderação for zero, negativo ou omissivo, o caso será excluído da análise.

## Métodos de Seleção de Variável de Regressão Linear

A seleção de método permite especificar como as variáveis independentes são inseridas na análise. Usando métodos diferentes, é possível construir uma variedade de modelos de regressão a partir do mesmo conjunto de variáveis.

- *Inserir (Regressão)*. Um procedimento para seleção de variáveis em que todas as variáveis em um bloco são inseridas em um único passo.
- *Stepwise*. Em cada etapa, a variável independente não na equação que tem a menor probabilidade de F é inserida, se essa probabilidade for suficientemente pequena. As variáveis que já estiverem na equação de regressão serão removidas se a probabilidade de F for suficientemente grande. O método finalizará quando não houver mais variáveis elegíveis para inclusão ou remoção.
- *Remove*. Um procedimento para seleção de variáveis em que todas as variáveis em um bloco são removidas em um único passo.
- *Eliminação Backward*. Um procedimento de seleção de variáveis no qual todas as variáveis são inseridas na equação, em seguida, removidas sequencialmente. A variável com a menor correlação parcial com a variável dependente é considerada primeiro para remoção. Se ela atender ao critério para eliminação, ela será removida. Após a primeira variável ser removida, a variável restante na equação com a menor correlação parcial é a próxima a ser considerada. O procedimento é interrompido quando não houver variáveis na equação que satisfaçam os critérios de remoção.
- *Seleção Forward*. Um procedimento de seleção de variáveis stepwise no qual as variáveis são inseridas sequencialmente no modelo. A primeira variável considerada para entrada na equação é aquela com a maior correlação positiva ou negativa com a variável dependente. Esta variável será inserida na equação somente se ela satisfizer os critérios para a entrada. Se a primeira variável for inserida, a variável independente que não estiver na equação e que possuir a maior correlação parcial é considerada a próxima. O procedimento é interrompido quando não houver variáveis que atendam ao critério de entrada.

Os valores de significância em sua saída baseiam-se no ajuste de um modelo único. Portanto, os valores de significância geralmente são inválidos quando um método stepwise (stepwise, forward ou backward) é usado.

Todas as variáveis devem passar pelo critério de tolerância para serem inseridas na equação, independentemente do método de entrada especificado. O nível de tolerância padrão é 0,0001. Além disso, uma variável não é inserida se fizer com que a tolerância de outra variável já no modelo fique abaixo do critério de tolerância.

Todas as variáveis independentes selecionadas são incluídas em um único modelo de regressão. Entretanto, é possível especificar métodos de entrada diferentes para subconjuntos diferentes de variáveis. Por exemplo, é possível inserir um bloco de variáveis no modelo de regressão utilizando a seleção stepwise e um segundo bloco utilizando a seleção forward. Para incluir um segundo bloco de variáveis no modelo de regressão, clique em **Avançar**.

## Regra do Conjunto de Regressão Linear

Os casos definidos pela regra de seleção são incluídos na análise. Por exemplo, se selecionar uma variável, escolher **igual a** e digitar 5 para o valor, então apenas os casos para os quais a variável selecionada tiver um valor igual a 5 serão incluídos na análise. Um valor da sequência de caracteres também é permitido.

## Gráficos de Regressão Linear

Os gráficos podem auxiliar na validação dos pressupostos de normalidade, linearidade e igualdade de variâncias. Os gráficos também são úteis para detectar discrepâncias, observações incomuns e casos influentes. Após salvá-los como novas variáveis, valores previstos, resíduos e outras informações de diagnóstico estão disponíveis no Editor de Dados para a construção de gráficos com as variáveis independentes. Os gráficos a seguir estão disponíveis:

**Gráficos de dispersão.** É possível representar quaisquer dois dos seguintes itens: a variável dependente, valores preditos padronizados, resíduos padronizados, resíduos excluídos, valores preditos ajustados, resíduos Estudentizados ou resíduos excluídos Estudentizados. Represente os resíduos padronizados com relação aos valores preditos padronizados para verificar a linearidade e a igualdade das variâncias.

*Lista de variáveis de origem.* Lista a variável dependente (DEPENDNT) e as variáveis preditas e residuais a seguir: Valores preditos padronizados (\*ZPRED), Resíduos padronizados (\*ZRESID), Resíduos excluídos (\*DRESID), Valores preditos ajustados (\*ADJPRED), Resíduos estudentizados (\*SRESID), Resíduos excluídos estudentizados (\*SDRESID).

**Produzir todos os gráficos parciais.** Exibe gráficos de dispersão dos resíduos de cada variável independente e os resíduos da variável dependente quando ambas as variáveis são regredidas separadamente no restante das variáveis independentes. Pelo menos duas variáveis independentes devem estar na equação para que um gráfico parcial seja produzido.

**Gráficos de Resíduos Padronizados.** É possível obter histogramas de resíduos padronizados e gráficos de probabilidade normal comparando a distribuição de resíduos padronizados com uma distribuição normal.

Se quaisquer gráficos forem solicitados, estatísticas básicas serão exibidas para os valores preditos padronizados e resíduos padronizados (\*DRESID e \*ZRESID).

## Regressão linear: salvando novas variáveis

É possível salvar valores previstos, resíduos e outras estatísticas úteis para informações de diagnóstico. Cada seleção adiciona uma ou mais novas variáveis ao seu arquivo de dados ativo.

**Valores Preditos.** Valores que o modelo de regressão prevê para cada caso.

- *Não padronizado.* O valor que o modelo prediz para a variável dependente.
- *Padronizado.* Uma transformação de cada valor previsto em sua forma padronizada. Ou seja, o valor predito médio é subtraído do valor predito, e a diferença é dividida pelo desvio padrão dos valores preditos. Os valores preditos padronizados possuem uma média de 0 e um desvio padrão de 1.
- *Ajustado.* O valor predito para um caso quando esse caso é excluído do cálculo dos coeficientes de regressão.
- *S.E. de predições significam.* Erros padrão dos valores preditos Uma estimativa do desvio padrão do valor médio da variável dependente para os casos que tiverem os mesmos valores das variáveis independentes.

**Distâncias.** Medidas para identificar casos com combinações incomuns de valores para as variáveis independentes e casos que podem ter grande impacto no modelo de regressão.

- *Mahalanobis.* Uma medida de quanto os valores de um caso nas variáveis independentes diferem da média de todos os casos.. Uma distância de Mahalanobis grande identifica um caso como tendo valores extremos em uma ou mais variáveis independentes.
- *Cook.* Uma medida de quanto os resíduos de todos os casos mudariam se um caso em particular fosse excluído do cálculo dos coeficientes de regressão. Um D de Cook grande indica que excluir um caso do cálculo das estatísticas de regressão altera os coeficientes substancialmente.
- *Valores de ponto de alavanca.* Medidas a influência de um ponto no ajuste da regressão. O ponto de alavanca centralizado varia de 0 (nenhuma influência sobre o ajuste) a  $(N-1)/N$ .

**Intervalos de Predição.** Os limites superior e inferior para intervalos de predição de média e individuais.

- *Média.* Limites inferior e superior (duas variáveis) para o intervalo de predição da resposta média predita.
- *Indivíduo.* Limites inferior e superior (duas variáveis) para o intervalo de predição da variável dependente para um caso único.
- *Intervalo de confiança.* Insira um valor entre 1 e 99.99 para especificar o nível de confiança para os dois Intervalos de Previsão.. Média ou Individual deve ser selecionado antes de inserir esse valor. Os valores típicos do intervalo de confiança são 90, 95, e 99.

**Residuais.** O valor real da variável dependente menos o valor previsto pela equação de regressão.

- *Não padronizado.* A diferença entre um valor observado e o valor previsto pelo modelo
- *Padronizado.* O resíduo dividido por uma estimativa de seu desvio padrão Resíduos padronizados, também conhecidos como resíduos de Pearson, possuem uma média de 0 e um desvio padrão de 1.
- *Estudentizado.* O residual dividido por uma estimativa de seu desvio-padrão que varia de caso para caso, dependendo da distância de valores de cada caso sobre as variáveis independentes a partir dos meios das variáveis independentes. Às vezes referido como resíduos estudados internamente.
- *Excluído.* O resíduo para um caso em que esse caso é excluído do cálculo dos coeficientes de regressão Ele é a diferença entre o valor da variável dependente e o valor previsto ajustado.
- *Studentized excluído.* O resíduo excluído para um caso dividido por seu erro padrão A diferença entre um residual deletado Studentizado e seu residual de Studentizado associado indica quanta diferença eliminando um caso faz em sua própria previsão. Às vezes referido como resíduos estudados externamente.

**Estatísticas de Influência.** A mudança nos coeficientes de regressão ( $DfBeta[s]$ ) e valores preditos ( $DfFit$ ) que resulta da exclusão de um caso particular. Valores padronizados de  $DfBetas$  e  $DfFit$  também estão disponíveis junto com a razão de covariância.

- *$DfBetas$ .* A diferença no valor beta é a mudança no coeficiente de regressão resultante da exclusão de um determinado caso. Um valor é calculado para cada termo no modelo, incluindo a constante.
- *$DfBeta$  Padronizado.* Diferença padronizada no valor beta. A mudança no coeficiente de regressão resultante da exclusão de um caso específico. Você pode querer analisar casos com valores absolutos maiores que 2 divididos pela raiz quadrada de  $N$ , em que  $N$  é o número de casos. Um valor é calculado para cada termo no modelo, incluindo a constante.
- *$DfFit$ .* A diferença no valor de ajuste é a mudança no valor previsto que resulta da exclusão de um caso específico.
- *$DfFit$  Padronizado.* Diferença padronizada no valor de ajuste A mudança no valor previsto resultante da exclusão de um caso específico. Você pode querer examinar valores padronizados que, em valor absoluto, excedem 2 vezes a raiz quadrada de  $p/N$ , em que  $p$  é o número paramétricas no modelo e  $N$  é o número de casos.
- *Razão de covariância.* A razão do determinante da matriz de covariâncias com um caso específico excluído do cálculo dos coeficiente de regressão para o determinante da matriz de covariâncias com todos os casos incluídos. Se a razão estiver próxima de 1, o caso não altera significativamente a matriz de covariâncias.

**Estatísticas de coeficiente.** Salva os coeficientes de regressão em um conjunto de dados ou em um arquivo de dados. Conjuntos de dados estão disponíveis para uso subsequente na mesma sessão, mas não são salvos como arquivos, a menos que sejam salvos explicitamente antes do término da sessão. Os nomes do conjunto de dados devem se adequar às regras de nomenclatura de variáveis. para obter mais informações.

**Exportar informações de modelo em arquivo XML.** As estimativas de parâmetro e (opcionalmente) suas covariâncias são exportadas no arquivo especificado em formato XML (PMML). É possível usar esse arquivo de modelo para aplicar as informações de modelo a outros arquivos de dados para propósitos de escoragem. para obter mais informações.

## Estatísticas de Regressão Linear

As seguintes estatísticas estão disponíveis:

**Coefficientes de Regressão- Estimativas** exibe o coeficiente de Regressão  $B$ , o erro padrão de  $B$ , o coeficiente beta padronizado, o valor  $t$  para  $B$  e o nível de significância bicaudal de  $t$ . Os **intervalos de confiança** exibem intervalos de confiança com o nível de confiança especificado para cada coeficiente de regressão ou uma matriz de covariância. **Matriz de covariâncias** exibe uma matriz de variâncias-covariâncias de coeficientes de regressão com covariâncias fora da diagonal e variâncias na diagonal. Uma matriz de correlações também é exibida.

**Ajuste de modelo-** As variáveis que são inseridas e removidas do modelo são listadas e as seguintes 'estatísticas de qualidade do ajuste' são exibidas: múltiplas  $R$ ,  $R^2$  e ajustadas  $R^2$ , erro padrão da estimativa e uma tabela de análise de variância.

**Mudança de R ao quadrado-** A mudança na estatística  $R^2$  que é produzida incluindo ou excluindo uma variável independente. Se a alteração de  $R^2$  associada a uma variável for grande, isso significa que a variável é uma boa preditora da variável dependente.

**Descritivo-** Fornece o número de casos válidos, a média e o desvio padrão para cada variável na análise.. Uma matriz de correlações com um nível de significância de um fator e o número de casos para cada correlação também são exibidos.

*Correlação Semiparcial.* A correlação entre a variável dependente e uma variável independente quando os efeitos lineares das outras variáveis independentes no modelo foram removidos da variável independente. Ela está relacionada à mudança no R-quadrado quando uma variável é incluída em uma equação. Às vezes é chamada de correlação semiparcial.

*Correlação Parcial.* A correlação que permanece entre duas variáveis após remover a correlação que é devido à sua associação mútua com outras variáveis. A correlação entre a variável dependente e uma variável independente quando os efeitos lineares das outras variáveis independentes no modelo tiverem sido removidos de ambas.

**Diagnóstico de Colinearidade-** Colinearidade (ou multicolinearidade) é a situação indesejável quando uma variável independente é uma função linear de outras variáveis independentes. Autovalores da matriz de produtos cruzados escalados e não centralizados, índices de condição e proporções de decomposição de variância são exibidos junto de fatores de inflação de variância (VIF) e tolerâncias para variáveis individuais.

**Crítérios de seleção-** Inclui o critério de informações de Akaike (AIC), o critério de predição de Ameniya (PC), o erro quadrático médio do critério de predição (Cp) e o critério Bayesiano de Schwarz (SBC). As estatísticas são exibidas na tabela de Resumo do Modelo

**Resíduos-** É possível selecionar '**PRESS Statistic**' para usar como uma estatística de validação cruzada para comparar diferentes modelos. Isso também exibe o teste '**Durban-Watson**' para correlação serial dos resíduos.. Escolha as informações de '**Diagnóstico de Caso**' para os casos que atendem ao critério de seleção (valores discrepantes acima de  $n$  desvios padrão).

## Opções de regressão linear

As opções a seguir estão disponíveis:

**Crítérios de Método Avançandos** Essas opções se aplicam quando o método de seleção de variáveis forward, backward ou stepwise é especificado. As variáveis podem ser inseridas ou removidas do modelo dependendo da significância (probabilidade) do valor  $F$  ou do próprio valor  $F$ .

- *Usar Probabilidade de  $F$ .* Uma variável será inserida no modelo se o nível de significância de seu valor  $F$  for menor que o valor de entrada e será removida se o nível de significância for maior que o valor de Remoção. A Entrada deve ser menor que Remoção, e ambos os valores devem ser positivos. Para inserir mais variáveis no modelo, aumente o valor de Entrada. Para remover mais variáveis do modelo, diminua o valor de Remoção.
- *Usar valor de  $F$ .* Uma variável será inserida no modelo se seu valor  $F$  for maior que o valor de Entrada e será removida se o valor  $F$  for menor que o valor de Remoção. A Entrada deve ser maior que Remoção, e ambos os valores devem ser positivos. Para inserir mais variáveis no modelo, diminua o valor de Entrada. Para remover mais variáveis do modelo, aumente o valor de Remoção.

**Tolerância.** Por padrão, o valor é 0,0001 Tolerância é a proporção da variância de uma variável na equação que não é considerada por outras variáveis independentes na equação. A tolerância mínima para qualquer variável na equação se a variável considerada foi incluída na análise é a tolerância mínima de uma variável que não está incluída na equação. As variáveis devem passar nos testes de tolerância e de tolerância mínima para entrar e permanecer em uma equação de regressão Se uma variável for aprovada nos critérios de tolerância, ela será elegível para inclusão com base no método em vigor

**Incluir constante na equação.** Por padrão, o modelo de regressão inclui um termo constante. Desmarcar esta opção força a regressão por meio da origem, o que raramente é feito. Alguns resultados de regressão por meio da origem não são comparáveis com os resultados de regressão que incluem uma constante. Por exemplo,  $R^2$  não pode ser interpretado de maneira habitual.

**Valores omissos.** É possível escolher uma das seguintes opções:

- **Excluir listwise dos casos.** Apenas casos com valores válidos para todas as variáveis são incluídos na análise.
- **Excluir casos entre pares.** Casos com dados completos para o par de variáveis que estão sendo correlacionadas são utilizados para calcular o coeficiente de correlação no qual a análise de regressão é baseada. Graus de liberdade são baseados no  $N$  mínimo entre pares.
- **Substituir pela média.** Todos os casos são utilizados para cálculos, com a média da variável substituída por observações omissas.

## Recursos adicionais do comando REGRESSION

O idioma da sintaxe de comando também permite:

- Gravar uma matriz de correlações ou ler uma matriz no lugar dos dados brutos para obter sua análise de regressão (com o subcomando MATRIX).
- Especificar níveis de tolerância (com o subcomando CRITERIA).
- Obter diversos modelos para as mesmas variáveis dependentes ou diferentes (com os subcomandos METHOD e DEPENDENT).
- Obter estatísticas adicionais (com os subcomandos DESCRIPTIVES e STATISTICS).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Regressão ordinal

---

A Regressão Ordinal lhe permite modelar a dependência de uma resposta ordinal politômica em um conjunto de preditores, que podem ser fatores ou covariáveis. O design de Regressão Ordinal é baseado na metodologia de McCullagh (1980, 1998) e o procedimento é referido como PLUM na sintaxe.

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 predictoras (independentes). Os coeficientes estimados refletem a maneira como as mudanças nos preditores afetam a resposta. A resposta é considerada como numérica, no sentido de que mudanças no nível da resposta são equivalentes em todo o intervalo da resposta. Por exemplo, a diferença entre uma pessoa de 1,50 m de altura e uma pessoa de 1,40 m de altura é de 10 cm, que tem o mesmo significado que a diferença entre uma pessoa de 2,10 m de altura e uma pessoa de 2 m de altura. Esses relacionamentos não necessariamente se mantêm para variáveis ordinais, em que a escolha e o número de categorias de resposta podem ser bastante arbitrários.

**Exemplo.** A Regressão Ordinal poderia ser usada para estudar a reação do paciente à dosagem de medicamentos. As possíveis reações podem ser classificadas como *nenhuma*, *suave*, *moderada* ou *grave*. A diferença entre uma reação suave e moderada é difícil ou impossível de quantificar e baseia-se na percepção. Além disso, a diferença entre uma resposta suave e moderada pode ser maior ou menor que a diferença entre uma resposta moderada e grave.

**Estatísticas e gráficos.** Frequências observadas e esperadas e frequências acumulativas, resíduos de Pearson para frequências e frequências acumulativas, probabilidades observadas e esperadas, probabilidades acumulativas observadas e esperadas de cada categoria de resposta por padrão de covariável, matrizes de correlação e de covariâncias assintóticas de estimativas paramétrica, qui-quadrado de Pearson e qui-quadrado de razão de verossimilhança, estatísticas de qualidade de ajuste, histórico de iteração, teste de suposição de linhas paralelas, estimativas dos parâmetro, erros padrão, intervalos de confiança e estatísticas  $R^2$  de Cox e Snell, de Nagelkerke e McFadden.

Considerações de Dados de Regressão Ordinal



**Dados.** A variável dependente é considerada como ordinal e pode ser numérica ou uma sequência de caracteres. A ordenação é determinada pela classificação dos valores da variável dependente em ordem crescente. O valor mais baixo define a primeira categoria. Variáveis de fator são consideradas categóricas. As variáveis de covariáveis devem ser numéricas. Observe que usar mais de uma covariável contínua pode facilmente resultar na criação de uma tabela de probabilidades de célula muito grande.

**Suposições.** Somente uma variável de resposta é permitida e ela deve ser especificada. Além disso, para cada padrão distinto de valores ao longo das variáveis independentes, as respostas são consideradas variáveis multinomiais independentes.

**Procedimentos relacionados.** A regressão logística nominal usa modelos semelhantes para variáveis dependentes nominais.

Obtendo uma Regressão Ordinal

1. A partir dos menus, escolha:  
**Analisar > Regressão > Ordinal...**
2. Selecione uma variável dependente.
3. Clique em **OK**.

## Opções de Regressão Ordinal

A caixa de diálogo Opções permite ajustar os parâmetros usados no algoritmo de estimativa iterativa, escolher um nível de confiança para suas estimativas de parâmetro e selecionar uma função de ligação.

**Iterações.** É possível customizar o algoritmo iterativo.

- **Máximo de iterações.** Especificar um número inteiro não negativo. Se 0 for especificado, o procedimento retorna as estimativas iniciais.
- **Máximo de redução para metade do passo.** Especifique um número inteiro positivo.
- **Convergência de log da verossimilhança.** O algoritmo para se a mudança absoluta ou relativa no log da verossimilhança for menor que esse valor. O critério não é usado se 0 for especificado.
- **Convergência de parâmetro.** O algoritmo para se a mudança absoluta ou relativa em cada uma das estimativas de parâmetro for menor que esse valor. O critério não é usado se 0 for especificado.

**Intervalo de confiança.** Especifique um valor maior ou igual a 0 e menor que 100.

**Delta.** O valor incluído em frequências de célula zero. Especifique um valor não negativo menor que 1.

**Tolerância à singularidade.** Usado para verificar se há preditores altamente dependentes. Selecione um valor a partir da lista de opções.

**Função de ligação.** A função de ligação é uma transformação das probabilidades cumulativas que permite estimar o modelo. As cinco funções de link a seguir estão disponíveis.

- **Logit.**  $f(x) = \log(x/(1-x))$ . Normalmente usada para categorias distribuídas uniformemente.
- **Log-log complementar.**  $f(x) = \log(-\log(1-x))$ . Normalmente usada quando as categorias mais altas são mais prováveis.
- **Log-log negativo.**  $f(x) = -\log(-\log(x))$ . Normalmente usada quando as categorias mais baixas são mais prováveis.
- **Probit.**  $f(x) = \delta^{-1}(x)$ . Normalmente usada quando a variável latente é normalmente distribuída.
- **Cauchit (inverse Cauchy).**  $f(x) = \tan(\pi(x-0.5))$ . Normalmente usada quando a variável latente tem muitos valores extremos.

## Saída de Regressão Ordinal

A caixa de diálogo Saída permite produzir tabelas para exibição no Visualizador e salvar variáveis no arquivo de trabalho.

**Exibição.** Produz tabelas para:

- **Imprimir histórico de iteração para cada n passo(s).** As estimativas de log da verossimilhança e paramétrica são impressas para a frequência de iteração de impressão especificada. A primeira e última iterações são sempre impressas.
- **Estatísticas de Qualidade do ajuste.** As estatísticas de Pearson e de qui-quadrado de razão de verossimilhança. Elas são calculadas com base na classificação especificada na lista de variáveis.
- **Estatística de sumarização.** Estatísticas  $R^2$  de Cox e Snell, de Nagelkerke e de McFadden.
- **Estimativas paramétrica.** Estimativas paramétricas, erros padrão e intervalos de confiança.
- **Correlação assintótica de estimativas paramétrica.** Matriz de correlações de estimativas paramétrica.
- **Covariância assintótica de estimativas paramétrica.** Matriz de covariâncias de estimativas de parâmetros.
- **Informações de célula.** Frequências observadas e esperadas e frequências acumulativas, resíduos de Pearson para frequências e frequências acumulativas, probabilidades observadas e esperadas, probabilidades acumulativas observadas e esperadas de cada categoria de resposta por padrão de covariável. Observe que para modelos com muitos padrões de covariável (por exemplo, modelos com covariáveis contínuas), essa opção pode gerar uma tabela muito grande e pesada.
- **Teste de linhas paralelas.** Teste da hipótese de que parâmetros de localização são equivalentes ao longo dos níveis da variável dependente. Isso está disponível somente para o modelo somente para local.

**Variáveis salvas.** Salva as variáveis a seguir no arquivo de trabalho:

- **Probabilidades de resposta estimada.** Probabilidades estimadas de modelo de classificação de um padrão de covariável/fator nas categorias de resposta. Há tantas probabilidades quanto o número de categorias de resposta.
- **Categoria predita.** A categoria de resposta que tem a probabilidade máxima estimada para um padrão de covariável/fator.
- **Probabilidade de categoria predita.** Probabilidade estimada de classificação de um padrão de covariável/fator na categoria predita. Essa probabilidade também é o máximo das probabilidades estimadas do padrão de covariável/fator.
- **Probabilidade de categoria real.** Probabilidade estimada de classificação de um padrão de covariável/fator na categoria real.

**Imprimir log da verossimilhança.** Controla a exibição do log da verossimilhança. **Incluir a constante de multinômio** lhe dá o valor integral da probabilidade. Para comparar os seus resultados ao longo de produtos que não incluem a constante, é possível escolher excluí-la.

## Modelo de Localização de Regressão Ordinal

A caixa de diálogo Localização permite especificar o modelo de localização para a sua análise.

**Especificar modelo.** Um modelo dos principais efeitos contém os efeitos principais de covariável e fator, mas nenhum efeito de interação. É possível criar um modelo customizado para especificar subconjuntos de interações de fatores ou interações de covariáveis.

**Fatores/covariáveis.** Os fatores e covariáveis são listados.

**Modelo de localização.** O modelo depende dos efeitos principais e dos efeitos de interação que você selecionar.

Para os fatores e covariáveis selecionados:

### Interação

Cria o termo de interação de mais alto nível para todas as variáveis selecionadas. Este é o padrão.

### Efeitos principais

Cria um termo dos principais efeitos para cada variável selecionada.

### Todos de 2 fatores

Cria todas as interações bidirecionais possíveis das variáveis selecionadas.

**Todos de 3 fatores**

Cria todas as interações de três vias possíveis das variáveis selecionadas.

**Todos de 4 fatores**

Cria todas as possíveis interações de quatro vias das variáveis selecionadas.

**Todos de 5 fatores**

Cria todas as possíveis interações de cinco vias das variáveis selecionadas.

**Termos de construção e termos customizados****termos de construção**

Use essa opção quando desejar incluir termos não aninhados de um determinado tipo (como efeitos principais) para todas as combinações de um conjunto selecionado de fatores e covariáveis.

**Construir termos customizados**

Use esta opção quando desejar incluir termos aninhados ou quando desejar construir explicitamente qualquer variável de termo por variável. A criação de um termo aninhado envolve as etapas a seguir:

**Modelo de Escala de Regressão Ordinal**

A caixa de diálogo Escala permite especificar o modelo de escala para a sua análise.

**Fatores/covariáveis.** Os fatores e covariáveis são listados.

**Modelo de escala.** O modelo depende dos efeitos principais e de interação que você selecionar.

Para os fatores e covariáveis selecionados:

**Interação**

Cria o termo de interação de mais alto nível para todas as variáveis selecionadas. Este é o padrão.

**Efeitos principais**

Cria um termo dos principais efeitos para cada variável selecionada.

**Todos de 2 fatores**

Cria todas as interações bidirecionais possíveis das variáveis selecionadas.

**Todos de 3 fatores**

Cria todas as interações de três vias possíveis das variáveis selecionadas.

**Todos de 4 fatores**

Cria todas as possíveis interações de quatro vias das variáveis selecionadas.

**Todos de 5 fatores**

Cria todas as possíveis interações de cinco vias das variáveis selecionadas.

**Termos de construção e termos customizados****termos de construção**

Use essa opção quando desejar incluir termos não aninhados de um determinado tipo (como efeitos principais) para todas as combinações de um conjunto selecionado de fatores e covariáveis.

**Construir termos customizados**

Use esta opção quando desejar incluir termos aninhados ou quando desejar construir explicitamente qualquer variável de termo por variável. A criação de um termo aninhado envolve as etapas a seguir:

**Recursos Adicionais do Comando PLUM**

É possível customizar sua Regressão Ordinal se colar suas seleções em uma janela de sintaxe e editar a sintaxe de comando PLUM resultante. O idioma da sintaxe de comando também permite:

- Crie testes de hipóteses customizados especificando hipóteses nulas como combinações lineares de parâmetros.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Regressão de rede elástica linear

Linear Elastic Net usa a classe Python `sklearn.linear_model.ElasticNet` para estimar modelos de regressão linear regularizados para uma variável dependente em uma ou mais variáveis independentes. A regularização combina penalidades L1 (Lasso) e L2 (Ridge). A extensão inclui modos opcionais para exibir gráficos de rastreamento para diferentes valores de alfa para uma determinada razão L1 e para selecionar a razão L1 e os valores de hiperparâmetro alfa com base na validação cruzada. Quando um modelo único é encaixado ou a validação cruzada é usada para selecionar a proporção de pênalti e / ou alfa, uma partição de dados de holdout pode ser usada para estimar o desempenho fora da amostra.

Além de adequar um modelo com valores especificados da proporção de L1 de pênaltis e parâmetro de regularização alfa, rede elástica linear pode exibir um traço de valores de coeficiente para uma gama de valores alfa para uma determinada proporção, ou facilitar a escolha do valor de hiperparâmetros via validação de cruzamento k-fold em grades especificadas de valores. Se um único modelo for ajustado ou a seleção de razão e / ou alfa por meio de validação cruzada for executada, o modelo final poderá ser aplicado aos dados mantidos que são criados por uma partição dos dados de entrada para obter uma estimativa válida do desempenho fora da amostra do modelo.

### Obtenção de uma análise de Regressão Líquida Elástica Linear

1. Nos menus, escolha:

**Analisar > Regressão > Alternativas de OLS Lineares > Elastic Net**

O diálogo permite especificar uma variável que designa cada caso no dataset ativo para a amostra de treinamento ou holdout.

2. Selecione uma variável de destino numérica. Somente uma variável de destino é necessária para executar uma análise.
3. Especificar um dependente numérico.
4. Especificar pelo menos uma variável de fator categórico ou variável de covariado numérico.

Opcionalmente, **Partition** fornece uma maneira de criar um subconjunto de holdout ou teste dos dados de entrada para estimativa de desempenho fora de amostra do modelo especificado ou escolhido. Todo o particionamento é realizado após a exclusão listwise de quaisquer casos com dados inválidos para qualquer variável utilizada pelo procedimento. Observe que para validação do crossvalidação, folds ou partições dos dados de treinamento são criados em Python. Os dados de validação criados pela partição não são usados na estimação, independentemente do modo em vigor.

A partição pode ser definida especificando a proporção de casos que são designados aleatoriamente a cada amostra (em **partições de treinamento e de suspensão**) ou por uma variável que designa cada caso à amostra de treinamento ou de validação. Não é possível especificar treinamentos e variáveis. Se a partição não for especificada, uma amostra holdout é criada de aproximadamente 30% dos dados de entrada é criada.

O **Treinamento%** especifica o número relativo de casos no dataset ativo para atribuir aleatoriamente a amostra de treinamento. O treinamento padrão é 70%.

### Regressão Líquida Elástica Lineares: Opções

A guia Opções fornece opções para:

#### Modo

Esta seleção fornece opções para especificar um dos modos a seguir:

#### Ajuste com proporção de L1 especificado e alfa

Quando você seleciona esse modo, um único modelo é ajustado para os dados de treinamento usando valores de regularização alfa e L1 especificados. Este é o padrão. Se uma partição for especificada, o modelo único ou final ajustado será aplicado aos dados de teste retidos para estimar o desempenho fora da amostra.

Em **Plotar**, gráficos de valores observados e / ou resíduos versus valores preditos podem ser selecionados

Em **Salvar**, você pode especificar valores previstos e resíduos para salvar.

### Gráfico de rastreo

Ao selecionar esse modo, três gráficos para os dados de treinamento são exibidos como uma função de alfa para o conjunto especificado de valores de alfa:

- Um traço traço de coeficientes de regressão.
- Um gráfico de  $R^2$ .
- Um complô de erro quadrado médio (MSE).

Apesar de a partição ser honrada, nenhum resultado para dados de teste realizado são fornecidos porque nenhum modelo final resulta deste modo.

### L1 razão e / ou seleção alfa por meio de validação cruzada

Ao selecionar esse modo, uma procura de grade com validação cruzada para avaliar modelos é executada e os melhores valores de razão e alfa são escolhidos com base na melhor média  $R^2$  sobre as dobras de validação. O campo **Número de dobras de validação cruzada** pode ser usado para alterar o valor padrão de cinco divisões ou dobras para validação cruzada. Se uma partição for especificada, o modelo único ou final ajustado será aplicado aos dados de teste retidos para estimar o desempenho fora da amostra.

Em **Exibir**, é possível selecionar para mostrar informações básicas sobre apenas o modelo com o valor escolhido de razão e alfa (**Melhor**), informações básicas sobre todos os modelos comparados (**Comparar modelos**) ou informações completas sobre todas as divisões ou dobras para todos os modelos (**Comparar modelos e dobras**). **Best** é o padrão.

Em **Plot**, gráficos da média  $R^2$  e / ou MSE sobre as dobras de validação podem ser selecionados. Também podem ser selecionados gráficos de valores observados e / ou de resíduos versus valores preditos

Em **Salvar**, você pode especificar valores previstos e resíduos para salvar.

### Especificar proporções L1 individuais

Ao selecionar esse modo para o modo **Ajustar com a razão L1 especificada e o modo alfa** ou **Gráfico de rastreo**, é possível especificar um único valor de proporção de penalidade L1. Quando selecionado para a proporção **L1 e / ou seleção alfa por meio do modo de validação cruzada**, é possível especificar diversos valores..

#### Especificar grade de valores alfa

Ao selecionar esse modo para o modo **L1 de proporção e / ou seleção alfa por meio da validação cruzada**, uma grade de valores alfa exclusivos pode ser especificada de um valor **Start** (value1) para um valor **End** (value2) com o incremento de **By** (value3). Se especificado, apenas um conjunto válido de [value1 TO value2 BY value3] será permitido. Ele deve satisfazer esse  $0 \leq \text{value1} \leq \text{value2} \leq 1$ . Nos casos em que  $\text{value1} = \text{value2}$ , isso é equivalente a especificar um único value1, independentemente de value3.

Os gráficos são exibidos usando a métrica especificada para os eixos X horizontais dos valores alfa variáveis..

### Especificar alphas individuais

Ao selecionar esse modo para o modo **Ajustar com a razão L1 especificada e o modo alfa**, é possível especificar um único valor de regularização alpha. Quando selecionado para o modo **Gráfico de rastreo** ou **L1 e / ou seleção alfa por meio de validação cruzada**, é possível especificar diversos valores.

#### Especificar grade de valores alfa

Quando você seleciona esse modo para o modo **Gráfico de Rastreo** ou **L1 e / ou seleção alfa por meio de validação cruzada**, uma grade de valores alfa exclusivos pode ser especificada de um valor **Start** (value1) para um valor **End** (value2) com o incremento de **By** (value3). Se especificado, apenas um conjunto válido de [value1 TO value2 BY value3] será permitido.

Deve satisfazer esse  $0 \leq \text{value1} \leq \text{value2} \leq 1$ . Nos casos em que  $\text{value1} = \text{value2}$ , é equivalente a especificação de um único  $\text{value1}$ , independentemente de  $\text{value3}$ .

A métrica **Alpha** para intervalos de valores pode ser **Linear** ou **Base 10 logarítmica** (10 é elevado à potência de valores especificados).

Os gráficos são exibidos usando a métrica especificada para os eixos X horizontais dos valores alfa variáveis..

## Critérios

Análises de controles.

### Incluir intercepto

Este critério inclui uma intercepção no (s) modelo (s) ajustado (s). Observe que o procedimento de extensão não centra ou padroniza a variável dependente, e a interceptação não é penalizada durante a estimação.

### Preditores padronizados

Padroniza todas as variáveis independentes.

### Número de dobras de validação cruzada

O número de divisões ou dobras para a avaliação de validação cruzada de modelos Deve ser um valor inteiro positivo maior que 1. O padrão é 5.

### Python estado aleatório

O valor da configuração `random_state` em Python usado enquanto você executa a avaliação de validação cruzada de modelos. Ele permite a replicação de resultados que envolvem números pseudo aleatórios. O valor deve ser um número inteiro no intervalo de 0 a  $2^{32}-1$ . O padrão é 0.

### Limite de tempo (minutos)

O número de minutos permitido para que os cálculos do modelo sejam executados. Se você especificar 0, o cronômetro é desligado. O valor padrão é 5 segundos.

## Exibir

Esta opção especifica a quantia de saída a ser exibida para o modo **L1 de proporção e / ou seleção alfa por meio de validação cruzada** .

### A(o) Melhor

Exibe apenas resultados básicos para o melhor modelo escolhido. Isso é configurado por padrão

### Comparar modelos

Exibe resultados básicos para todos os modelos avaliados.

### Comparar modelos e pastas

Exibe resultados de verbose completos para cada divisão ou dobra para cada modelo avaliado.

## Gráfico

Esta opção especifica gráficos de valores observados ou residuais versus valores preditos e, com validação cruzada, especificação de gráficos de erro quadrático médio (MSE) e / ou  $R^2$  sobre dobras de validação cruzada versus valores alfa.

### Média de erro quadrático médio de validação cruzada (MSE) versus alfa

Para a **L1 razão e / ou seleção alfa via validação cruzada** modo, exibe um gráfico de linha de média MSE sobre dobras de validação cruzada versus alfa para o especificado ou selecionado melhor L1 valor de razão. Para o modo **Trace plot** , uma parcela similar é produzida automaticamente com base nos dados completos de treinamento.

### Média de validação cruzada R Square versus alfa

Para a **L1 razão e / ou seleção alfa por meio de validação cruzada** modo, exibe um gráfico de linha de média  $R^2$  sobre dobras de validação cruzada versus alfa para o melhor valor de razão L1 especificado ou selecionado. Para o modo **Trace plot** , uma parcela similar é produzida automaticamente com base nos dados completos de treinamento.

### Observado versus Previsto

Exibe um gráfico de dispersão de valores observados em relação aos preditos para o modelo especificado ou melhor.

### Resíduos versus Previsto

Exibe um gráfico de dispersão de valores residuais em relação aos preditos para o modelo especificado ou melhor.

### Salvar

Especifica variáveis para salvar no conjunto de dados ativos.

### Valores preditos

Economiza valores previstos do modelo especificado ou melhor para o conjunto de dados ativos. Também é possível especificar um **Nome de variável customizada**

### Residuais

Guarde resíduos das previsões de modelo especificado ou melhor para o conjunto de dados ativos. Também é possível especificar um **Nome de variável customizada**

## Regressão linear lasso

---

O Lasso linear usa a classe Python `sklearn.linear_model.Lasso` para estimar modelos de regressão linear regularizados de perda L1 para uma variável dependente em uma ou mais variáveis independentes e inclui modos opcionais para exibir gráficos de rastreamento e selecionar o valor de hiperparâmetro alfa com base na validação cruzada. Quando um modelo único é montado ou a validação cruzada é usada para selecionar alfa, uma partição de dados de holdout pode ser usada para estimar o desempenho fora da amostra.

Além de encaixar um modelo com um valor especificado do parâmetro de regularização alfa, lasso linear pode exibir um traço de valores de coeficiente para uma gama de valores alfa, ou facilitar a escolha do valor de hiperparâmetro via crossvalidação de k-fold em grades especificadas de valores. Se um único modelo for ajustado ou a seleção alfa por meio de validação cruzada for executada, o modelo final poderá ser aplicado aos dados retidos criados por uma partição dos dados de entrada para obter uma estimativa válida do desempenho fora da amostra do modelo.

### Obtendo uma análise de Lasso Regressão Linear

1. Nos menus, escolha:

**Analisar > Regressão > Alternativas de OLS Lineares > Lasso**

O diálogo permite especificar uma variável que designa cada caso no dataset ativo para a amostra de treinamento ou holdout.

2. Selecione uma variável de destino numérica. Somente uma variável de destino é necessária para executar uma análise.

3. Especificar um dependente numérico.

4. Especificar pelo menos uma variável de fator categórico ou variável de covariado numérico.

Opcionalmente, **Partition** fornece uma maneira de criar um subconjunto de holdout ou teste dos dados de entrada para estimativa de desempenho fora de amostra do modelo especificado ou escolhido. Todo o particionamento é realizado após a exclusão listwise de quaisquer casos com dados inválidos para qualquer variável utilizada pelo procedimento. Observe que para validação do crossvalidação, folds ou partições dos dados de treinamento são criados em Python. Os dados de holdout criados pela partição não são usados em estimação, independentemente do modo em vigor.

A partição pode ser definida especificando a proporção de casos aleatoriamente atribuídos a cada amostra (sob **Partituras de Treinamento e Holdout**), ou por uma variável que atribui cada caso à amostra de treinamento ou holdout. Não é possível especificar treinamentos e variáveis. Se a partição não for especificada, uma amostra holdout é criada de aproximadamente 30% dos dados de entrada é criada.

O **Treinamento%** especifica o número relativo de casos no dataset ativo para atribuir aleatoriamente a amostra de treinamento. O treinamento padrão é 70%.

## Regressão Linear Lasso: Opções

A guia Opções fornece opções para:

### Modo

Esta seleção fornece opções para especificar um dos modos a seguir:

#### Ajuste com alfa especificado

Ao selecionar esse modo, um único modelo é ajustado aos dados de treinamento usando apenas um valor de regularização alfa. Isso é configurado por padrão. Se uma partição for especificada, o modelo único ou final ajustado será aplicado aos dados de teste retidos para estimar o desempenho fora da amostra.

Em **Plotar**, gráficos de valores observados e / ou resíduos versus valores preditos podem ser selecionados

Em **Salvar**, você pode especificar valores previstos e resíduos para salvar.

#### Gráfico de rastreo

Ao selecionar esse modo, três gráficos para os dados de treinamento são exibidos como uma função de alfa para o conjunto especificado de valores de alfa:

- Um traço de coeficientes de regressão.
- Um gráfico de  $R^2$ .
- Um complô de erro quadrado médio (MSE).

Apesar de a partição ser honrada, nenhum resultado para dados de teste realizado são fornecidos porque nenhum modelo final resulta deste modo.

#### Seleção alfa por meio de validação cruzada

Ao selecionar uma procura de grade com validação cruzada para avaliar modelos e escolher o melhor alfa com base na melhor média,  $R^2$  sobre as dobras de validação. O campo **Número de dobras de validação cruzada** pode ser usado para alterar o valor padrão de cinco divisões ou dobras para validação cruzada. Se uma partição for especificada, o modelo único ou final ajustado será aplicado aos dados de teste retidos para estimar o desempenho fora da amostra.

Em **Exibir**, é possível selecionar para mostrar informações básicas sobre apenas o modelo com o valor escolhido de alfa (**Melhor**), informações básicas sobre todos os modelos comparados (**Comparar modelos**) ou informações completas sobre todas as divisões ou dobras para todos os modelos (**Comparar modelos e dobras**). **Best** é o padrão.

Em **Plot**, gráficos da média  $R^2$  e / ou MSE sobre as dobras de validação podem ser selecionados. Também podem ser selecionados gráficos de valores observados e / ou de resíduos versus valores preditos

Em **Salvar**, você pode especificar valores previstos e resíduos para salvar.

#### Especificar alphas individuais

Ao selecionar o modo **Ajustar com alpha especificado**, é possível especificar um único valor de regularização alpha. Ao selecionar o modo **Gráfico de rastreo** ou **Seleção alfa por meio de validação cruzada**, é possível especificar diversos valores.

#### Valor(es)

Especificar um ou mais valores positivos de regularização alfa. Vários valores podem ser especificados individualmente ou como intervalos. O padrão é 1.

#### Especificar grade de valores alfa

Ao selecionar o modo **Gráfico de rastreo** ou **Seleção alfa por meio de validação cruzada**, uma grade de valores alfa exclusivos pode ser especificada a partir de um valor de **Início** (value1) para



um valor de **Término** (value2) com o incremento de **Por** (value3). Se especificado, apenas um conjunto válido de [value1 TO value2 BY value3] será permitido. Ele deve satisfazer esse  $0 \leq \text{value1} \leq \text{value2} \leq 1$ . Nos casos em que  $\text{value1} = \text{value2}$ , isso é equivalente a especificar um único value1, independentemente de value3.

A métrica **Alpha** para intervalos de valores pode ser **Linear** ou **Base 10 logarítmica** (10 é elevado à potência de valores especificados).

Os gráficos são exibidos usando a métrica especificada para os eixos X horizontais dos valores alfa variáveis..

## Critérios

Análises de controles.

### Incluir intercepto

Inclui uma interceptação no (s) modelo (s) montado (s). Observe que o procedimento de extensão não centra ou padroniza a variável dependente, e a interceptação não é penalizada durante a estimação.

### Preditores padronizados

Padroniza todas as variáveis independentes.

### Número de dobras de validação cruzada

O número de divisões ou dobras para a avaliação de validação cruzada de modelos Deve ser um valor inteiro positivo maior que 1. O padrão é 5.

### Python estado aleatório

O valor da configuração random\_state em Python usado ao executar a avaliação de validação cruzada de modelos. Permite a replicação de resultados que envolvem números pseudo aleatórios. Deve ser um número inteiro na faixa de  $0$  a  $2^{32}-1$ . O padrão é 0.

### Limite de tempo (minutos)

O número de minutos permitido para que os cálculos do modelo sejam executados. Se você especificar 0, o cronômetro é desligado. O valor padrão é 5 segundos.

## Exibir

Especifica a quantidade de saída a ser exibida para o modo **Seleção Alpha via validação cruzada** .

### A(o) Melhor

Exibe apenas resultados básicos para o melhor modelo escolhido. Este é o padrão.

### Comparar modelos

Exibe resultados básicos para todos os modelos avaliados.

### Comparar modelos e pastas

Exibe resultados de verbose completos para cada divisão ou dobra para cada modelo avaliado.

## Gráfico

Especifica gráficos de valores observados ou residuais versus valores preditos e, com validação cruzada, especificação de gráficos de erro quadrático médio médio (MSE) e / ou R médio<sup>2</sup> sobre dobras de validação cruzada versus valores alfa.

### Média de erro quadrático médio de validação cruzada (MSE) versus alfa

Para o modo **Seleção Alfa por meio de validação cruzada** , exibe um gráfico de linha de média MSE sobre dobras de validação cruzada versus alfa. Para o modo **Trace plot** , uma parcela similar é produzida automaticamente com base nos dados completos de treinamento.

### Média de validação cruzada R Square versus alfa

Para o modo **Seleção Alfa por meio de validação cruzada** , exibe um gráfico de linha da média R<sup>2</sup> sobre dobras de validação cruzada versus alfa. Para o modo **Trace plot** , uma parcela similar é produzida automaticamente com base nos dados completos de treinamento.

### **Observado versus Previsto**

Exibe um gráfico de dispersão de valores observados em relação aos preditos para o modelo especificado ou melhor.

### **Resíduos versus Previsto**

Exibe um gráfico de dispersão de valores residuais em relação aos preditos para o modelo especificado ou melhor.

### **Salvar**

Especifica variáveis para salvar no conjunto de dados ativos.

### **Valores preditos**

Economiza valores previstos do modelo especificado ou melhor para o conjunto de dados ativos.

Também é possível especificar um **Nome de variável customizada**

### **Residuais**

Guarde resíduos das previsões de modelo especificado ou melhor para o conjunto de dados ativos.

Também é possível especificar um **Nome de variável customizada**

## **Regressão de Ridge linear**

---

O Linear Ridge usa a classe Python `sklearn.linear_model.Ridge` para estimar L2 ou modelos de regressão linear regularizados de perda quadrática para uma variável dependente em uma ou mais variáveis independentes e inclui modos opcionais para exibir gráficos de rastreamento e para selecionar o valor de hiperparâmetro alfa com base na validação cruzada. Quando um modelo único é montado ou a validação cruzada é usada para selecionar alfa, uma partição de dados de holdout pode ser usada para estimar o desempenho fora da amostra.

Além de encaixar um modelo com um valor especificado do parâmetro de regularização alfa, o ridge linear pode exibir um traço de rastro de valores de coeficientes para uma gama de valores alfa, ou facilitar a escolha do valor de hiperparâmetro via crossvalidação de k-fold em grades especificadas de valores. Se um único modelo for ajustado ou a seleção alfa por meio de validação cruzada for executada, o modelo final poderá ser aplicado aos dados retidos criados por uma partição dos dados de entrada para obter uma estimativa válida do desempenho fora da amostra do modelo.

### **Obtendo uma análise de regressão de Linear Ridge Regression**

1. Nos menus, escolha:

**Analisar > Regressão > Alternativas de OLS Lineares > Ridge**

O diálogo permite especificar uma variável que designa cada caso no dataset ativo para a amostra de treinamento ou holdout.

2. Selecione uma variável de destino numérica. Somente uma variável de destino é necessária para executar uma análise.

3. Especificar um dependente numérico.

4. Especificar pelo menos uma variável de fator categórico ou variável de covariado numérico.

Opcionalmente, **Partition** fornece uma maneira de criar um subconjunto de holdout ou teste dos dados de entrada para estimativa de desempenho fora de amostra do modelo especificado ou escolhido. Todo o particionamento é realizado após a exclusão listwise de quaisquer casos com dados inválidos para qualquer variável utilizada pelo procedimento. Observe que para validação do crossvalidação, folds ou partições dos dados de treinamento são criados em Python. Os dados de validação criados pela partição não são usados na estimação, independentemente do modo em vigor.

A partição pode ser definida especificando a proporção de casos que são designados aleatoriamente a cada amostra (em **partições de treinamento e de suspensão**) ou por uma variável que designa cada caso à amostra de treinamento ou de validação. Não é possível especificar treinamentos e variáveis. Se a partição não for especificada, uma amostra holdout é criada de aproximadamente 30% dos dados de entrada é criada.

O **Treinamento%** especifica o número relativo de casos no dataset ativo para atribuir aleatoriamente a amostra de treinamento. O treinamento padrão é 70%.

## Regressão Linear Ridge: Opções

A guia Opções fornece opções para:

### Modo

Esta seleção fornece opções para especificar um dos modos a seguir:

#### Ajuste com alfa especificado

Ao selecionar essa opção, um único modelo é ajustado aos dados de treinamento que usam apenas um valor de regularização alfa. Isso é configurado por padrão. Se uma partição for especificada, o modelo único ou final ajustado será aplicado aos dados de teste retidos para estimar o desempenho fora da amostra.

Em **Plotar**, gráficos de valores observados e / ou resíduos versus valores preditos podem ser selecionados

Em **Salvar**, você pode especificar valores previstos e resíduos para salvar.

#### Gráfico de rastreo

Ao selecionar essa opção, três gráficos para os dados de treinamento são exibidos como uma função de alfa para o conjunto especificado de valores de alfa:

- Um traço de rastro de coeficientes de regressão.
- Um gráfico de  $R^2$ .
- Um complô de erro quadrado médio (MSE).

Apesar de a partição ser honrada, nenhum resultado para dados de teste realizado são fornecidos porque nenhum modelo final resulta deste modo.

#### Seleção alfa por meio de validação cruzada

Ao selecionar essa opção, uma procura de grade com validação cruzada para avaliar modelos é feita e o melhor alfa é escolhido com base na melhor média  $R^2$  sobre as dobras de validação. O campo **Número de dobras de validação cruzada** pode ser usado para alterar o valor padrão de cinco divisões ou dobras para validação cruzada. Se uma partição for especificada, o modelo único ou final ajustado será aplicado aos dados de teste retidos para estimar o desempenho fora da amostra.

Em **Exibir**, é possível selecionar para mostrar informações básicas sobre apenas o modelo com o valor escolhido de alfa (**Melhor**), informações básicas sobre todos os modelos comparados (**Comparar modelos**) ou informações completas sobre todas as divisões ou dobras para todos os modelos (**Comparar modelos e dobras**). **Best** é o padrão.

Em **Plot**, gráficos da média  $R^2$  e / ou MSE sobre as dobras de validação podem ser selecionados. Também podem ser selecionados gráficos de valores observados e / ou de resíduos versus valores preditos

Em **Salvar**, você pode especificar valores previstos e resíduos para salvar.

#### Especificar alphas individuais

Ao selecionar o modo **Ajustar com alpha especificado**, é possível especificar um único valor de regularização alpha. Ao selecionar o modo **Gráfico de rastreo** ou **Seleção alfa por meio de validação cruzada**, é possível especificar diversos valores.

#### Valor(es)

Especificar um ou mais valores positivos de regularização alfa. É possível especificar diversos valores individualmente, ou como intervalos,. O padrão é 1.

#### Especificar grade de valores alfa

Ao selecionar o modo **Gráfico de rastreo** ou **Seleção alfa por meio de validação cruzada**, uma grade de valores alfa exclusivos pode ser especificada a partir de um valor de **Início** (value1) para

um valor de **Término** (value2) com o incremento de **Por** (value3). Se especificado, apenas um conjunto válido de [value1 TO value2 BY value3] será permitido. Ele deve satisfazer esse  $0 \leq \text{value1} \leq \text{value2} \leq 1$ . Nos casos em que  $\text{value1} = \text{value2}$ , isso é equivalente a especificar um único value1, independentemente de value3.

A métrica **Alpha** para intervalos de valores pode ser **Linear** ou **Base 10 logarítmica** (10 é elevado à potência de valores especificados).

Os gráficos são exibidos usando a métrica especificada para os eixos X horizontais dos valores alfa variáveis..

## Critérios

Análises de controles.

### Incluir intercepto

Este critério inclui uma interceptação num ou mais modelos ajustados. Observe que o procedimento de extensão não centra ou padroniza a variável dependente, e a interceptação não é penalizada durante a estimação.

### Preditores padronizados

Esse critério padroniza todas as variáveis independentes..

### Número de dobras de validação cruzada

Use esses critérios para configurar o número de divisões ou dobras para a avaliação de validação cruzada de modelos. O número deve ser um valor de número inteiro positivo maior que 1. A configuração padrão é 5.

### Python estado aleatório

O valor da configuração random\_state em Python é usado ao executar a avaliação de validação cruzada de modelos. Permite a replicação de resultados que envolvem números pseudo aleatórios. Deve ser um número inteiro na faixa de  $0$  a  $2^{32}-1$ . O padrão é 0.

### Limite de tempo (minutos)

O número de minutos permitido para que os cálculos do modelo sejam executados. Se você especificar 0, o cronômetro é desligado. O valor padrão é 5 segundos.

## Exibir

Esta seção especifica a quantidade de saída a ser exibida para o modo **Seleção Alfa via validação cruzada**.

### A(o) Melhor

Ele exibe apenas os resultados básicos para o melhor modelo escolhido. Isso é configurado por padrão.

### Comparar modelos

Também exibe resultados básicos para todos os modelos avaliados.

### Comparar modelos e pastas

Por fim, ele exibe resultados detalhados completos para cada divisão ou dobra para cada modelo avaliado.

## Gráfico

Especifica gráficos de valores observados ou residuais versus valores preditos, com validação cruzada, especificação de gráficos de erro quadrático médio (MSE) e / ou R médio<sup>2</sup> sobre dobras de validação cruzada versus valores alfa.

### Média de erro quadrático médio de validação cruzada (MSE) versus alfa

O modo **Seleção Alfa por meio de validação cruzada** exibe um gráfico de linha de média MSE sobre dobras de validação cruzada versus alfa. Para o modo **Trace plot**, uma parcela similar é produzida automaticamente com base nos dados completos de treinamento.

### Média de validação cruzada R Square versus alfa

O modo **Seleção alfa por meio de validação cruzada** exibe um gráfico de linha da média  $R^2$  sobre dobras de validação cruzada versus alfa. Para o modo **Trace plot**, uma parcela similar é produzida automaticamente com base nos dados completos de treinamento.

### Observado versus Previsto

Ele exibe um gráfico de dispersão de valores observados versus preditos para o modelo especificado ou melhor.

### Resíduos versus Previsto

Também exibe um gráfico de dispersão de resíduos versus valores preditos para o modelo especificado ou melhor.

## Salvar

Especifica variáveis para salvar no conjunto de dados ativos.

### Valores preditos

Economiza valores previstos do modelo especificado ou melhor para o conjunto de dados ativos. Caso contrário, é possível especificar um **Nome de variável customizada**.

### Residuais

Guarde resíduos das previsões de modelo especificado ou melhor para o conjunto de dados ativos. Caso contrário, é possível especificar um **Nome de variável customizada**.

## Estimativa de curva

---

O procedimento Curva de Estimção produz estatísticas de regressão de curva de estimção e gráficos relacionados para 11 modelos de regressão de curva de estimção diferentes. Um modelo separado é produzido para cada variável dependente. Também é possível salvar valores preditos, resíduos e intervalos de predição como novas variáveis.

**Exemplo.** Um provedor de serviços da Internet rastreia a porcentagem de tráfego de e-mail infectado por vírus em suas redes ao longo do tempo. Um gráfico de dispersão revela que a relação não é linear. Você pode ajustar um modelo quadrático ou cúbico aos dados e verificar a validade das suposições e a qualidade do ajuste do modelo.

**Estatísticas.** Para cada modelo: coeficientes de regressão,  $R$  múltiplo,  $R^2$ ,  $R^2$  ajustado, erro padrão da estimativa, tabela de análise de variância, valores preditos, resíduos e intervalos de predição. Modelos: linear, logarítmico, inverso, quadrático, cúbico, potência, composto, curva S, logístico, crescimento e exponencial.

Considerações sobre dados de estimativa de curva

**Dados.** As variáveis dependentes e independentes devem ser quantitativas. Se selecionar **Tempo** a partir do conjunto de dados ativo como a variável independente (ao invés de selecionar uma variável), o procedimento Curva de Estimção gera uma variável de tempo em que o período de tempo entre os casos é uniforme. Se **Tempo** for selecionado, a variável dependente deverá ser uma medida de séries temporais. A análise de séries temporais requer uma estrutura de arquivo de dados na qual cada caso (linha) representa um conjunto de observações em um tempo diferente, e o período de tempo entre os casos é uniforme.

**Suposições.** Analise seus dados graficamente para determinar como as variáveis independentes e dependentes estão relacionadas (linearmente, exponencialmente, etc.). Os resíduos de um bom modelo devem ser distribuídos aleatoriamente e normais. Se for usado um modelo linear, as seguintes premissas devem ser atendidas: Para cada valor da variável independente, a distribuição da variável dependente deve ser normal. A variância da distribuição da variável dependente deve ser constante para todos os valores da variável independente. A relação entre a variável dependente e a variável independente deve ser linear e todas as observações devem ser independentes.

Para obter uma estimativa de curva

1. A partir dos menus, escolha:

### Analisar > Regressão > Curva de Estimação...

2. Selecione uma ou mais variáveis dependentes. Um modelo separado é produzido para cada variável dependente.
3. Selecione uma variável independente (ou selecione uma variável no conjunto de dados ativo ou selecione **Tempo**).
4. Opcionalmente:
  - Selecione uma variável para rotular casos em gráficos de dispersão. Para cada ponto no gráfico de dispersão, você pode usar a ferramenta Seleção de ponto para exibir o valor da variável Rótulo de caso.
  - Clique em **Salvar** para salvar valores preditos, resíduos e intervalos de predição como novas variáveis.

As opções a seguir também estão disponíveis:

- **Incluir constante na equação.** Estima um termo constante na equação de regressão. A constante é incluída por padrão.
- **Modelos de gráfico.** Representa os valores da variável dependente e cada modelo selecionado com relação à variável independente. Um gráfico separado é produzido para cada variável dependente.
- **Exibir tabela ANOVA.** Exibe uma tabela de análise de variância de resumo para cada modelo selecionado.

## Modelos de Curva de Estimação

É possível escolher um ou mais modelos de regressão de estimação de curva. Para determinar qual modelo usar, plote seus dados. Se as suas variáveis parecem estar relacionadas linearmente, use um modelo de regressão linear simples. Quando suas variáveis não estão relacionadas linearmente, tente transformar seus dados. Quando uma transformação não ajuda, você pode precisar de um modelo mais complicado. Visualize um gráfico de dispersão de seus dados; se o gráfico se assemelhar a uma função matemática que você reconhece, ajuste seus dados a esse tipo de modelo. Por exemplo, se seus dados se assemelham a uma função exponencial, use um modelo exponencial.

*Linear.* Modelo cuja equação é  $Y = b_0 + (b_1 * t)$ . Os valores de série são modelados como uma função linear de tempo.

*Logarítmico.* Modelo cuja equação é  $Y = b_0 + (b_1 * \ln(t))$ .

*Inverso.* Modelo cuja equação é  $Y = b_0 + (b_1 / t)$ .

*Quadrático.* Modelo cuja equação é  $Y = b_0 + (b_1 * t) + (b_2 * t^{**2})$ . O modelo quadrático pode ser usado para modelar uma série que "decola" ou uma série que amortece.

*Cúbico.* Modelo que é definido pela equação  $Y = b_0 + (b_1 * t) + (b_2 * t^{**2}) + (b_3 * t^{**3})$ .

*Potência.* Modelo cuja equação é  $Y = b_0 * (t^{**b_1})$  ou  $\ln(Y) = \ln(b_0) + (b_1 * \ln(t))$ .

*Composto.* Modelo cuja equação é  $Y = b_0 * (b_1^{**t})$  ou  $\ln(Y) = \ln(b_0) + (\ln(b_1) * t)$ .

*Curva S.* Modelo cuja equação é  $Y = e^{**}(b_0 + (b_1/t))$  ou  $\ln(Y) = b_0 + (b_1/t)$ .

*Logístico.* Modelo cuja equação é  $Y = 1 / (1/u + (b_0 * (b_1 * t)))$  ou  $\ln(1/y-1/u) = \ln(b_0) + (\ln(b_1)*t)$  em que u é o valor limite superior. Após selecionar Logística, especifique o valor limite superior a ser usado na equação de regressão. O valor deve ser um número positivo maior que o maior valor de variável dependente.

*Crescimento.* Modelo cuja equação é  $Y = e^{**}(b_0 + (b_1 * t))$  ou  $\ln(Y) = b_0 + (b_1 * t)$ .

*Exponencial.* Modelo cuja equação é  $Y = b_0 * (e^{**}(b_1 * t))$  ou  $\ln(Y) = \ln(b_0) + (b_1 * t)$ .

## Salvar na Curva de Estimação

**Salvar Variáveis.** Para cada modelo selecionado, é possível salvar valores preditos, resíduos (valor observado da variável dependente menos o valor predito do modelo) e intervalos de predição (limites superior e inferior). Os novos nomes de variáveis e rótulos descritivos são exibidos em uma tabela na janela de saída.

**Prever Casos.** No conjunto de dados ativo, se selecionar **Tempo** ao invés de uma variável como a variável independente, será possível especificar um período de previsão além do término da série temporal. É possível escolher uma das seguintes alternativas:

- **Prever a partir do período de estimação até o último caso.** Prevê valores para todos os casos no arquivo, com base nos casos no período de estimação. O período de estimativa, exibido na parte inferior da caixa de diálogo, é definido com a subcaixa de diálogo Intervalo da opção Selecionar Casos no menu Dados. Se nenhum período de estimativa foi definido, todos os casos são usados para prever valores.
- **Prever através de.** Prediz valores por meio da data, hora ou número de observação especificado, com base nos casos no período de estimação. Esse recurso pode ser usado para prever valores além do último caso na série temporal. As variáveis de data definidas atualmente determinam quais caixas de texto estão disponíveis para especificar o final do período de previsão. Se não houver variáveis de data definidas, é possível especificar o número de observação final (caso).

Use a opção Definir datas no menu Dados para criar variáveis de data.

## Regressão por quadrados mínimos parciais

O procedimento Regressão por Quadrados Mínimos Parciais estima modelos de regressão de quadrados mínimos parciais (PLS, também conhecido como "projeção para estrutura latente"). O PLS é uma técnica preditiva que é uma alternativa para a regressão por quadrados mínimos (OLS) ordinária, correlação canônica ou modelagem de equação estrutural, e é particularmente útil quando variáveis preditoras estão altamente correlacionadas ou quando o número de preditores excede o número de casos.

O PLS combina os recursos de análise de componentes principais e de diversas regressões. Ele primeiro extrai um conjunto de fatores latentes que explicam o máximo de covariância possível entre as variáveis independentes e dependentes. Em seguida, um passo de regressão prevê valores das variáveis dependentes utilizando a decomposição das variáveis independentes.

### Tabelas

Proporção da variância explicada (por fator latente), ponderações de fatores latentes, carregamentos de fatores latentes, a importância de variável independente na projeção (VIP) e as estimativas de parâmetros de regressão (por variável dependente) são todos produzidos por padrão.

### Gráficos

Importância de variável na projeção (VIP), escores de fatores, ponderações de fatores para os três primeiros fatores latentes e a distância para o modelo são todos produzidos a partir da guia [Opções](#).

## Considerações de dados

### Nível de medição

Variáveis dependentes e independentes (preditoras) podem ser de escala, nominais ou ordinais. O procedimento supõe que o nível de medição apropriado foi designado para todas as variáveis, embora seja possível alterar temporariamente o nível de medição para uma variável clicando com o botão direito na variável na lista de variáveis de origem e selecionando um nível de medição no menu pop-up. Variáveis categóricas (nominais ou ordinais) são tratadas de maneira equivalente pelo procedimento.

### Codificação de variável categórica

O procedimento recodifica temporariamente variáveis dependentes categóricas utilizando uma codificação um-de-c durante o procedimento. Se houver as categorias  $c$  de uma variável, então a variável é armazenada como vetores  $c$ , com a primeira categoria denotada  $(1, 0, \dots, 0)$ , a próxima categoria  $(0, 1, 0, \dots, 0)$ , ..., e a categoria final  $(0, 0, \dots, 0, 1)$ . Variáveis dependentes categóricas são representadas utilizando codificação simulada, isto é, simplesmente omite o indicador correspondente para a categoria de referência.

### Ponderações de frequência

Valores de ponderação são arredondados para o número inteiro mais próximo antes do uso. Casos com ponderações omissas ou ponderações menores que 0,5 não são utilizados na análise.

### Valores omissos

Os valores omissos do usuário e do sistema são tratados como inválidos.

## Ajustando novamente a escala

Todas as variáveis de modelo são centralizadas e padronizadas, incluindo variáveis indicadoras representando as variáveis categóricas.

## Obtendo a regressão parcial por quadrados mínimos

Nos menus, escolha:

### **Analisar > Regressão > Quadrados Mínimos Parciais ...**

1. Selecione pelo menos uma variável dependente.
2. Selecione pelo menos uma variável independente.

Como opção, você pode:

- Especificar uma categoria de referência para variáveis dependentes categóricas (nominais ou ordinais).
- Especificar uma variável a ser utilizada como um identificador exclusivo para a saída de casos e conjuntos de dados salvos.
- Especificar um limite superior no número de fatores latentes a serem extraídos.

## Pré-requisitos

O Procedimento de Reregressão de Squares Menos Squares é um comando de extensão Python e requer a funcionalidade Python, que é uma parte do seu produto IBM SPSS Statistics. Ele requer também as bibliotecas Python NumPy e SciPy, que estão gratuitamente disponíveis.

**Nota:** Para os usuários que trabalham no modo de análise distribuída (requer o IBM SPSS Statistics Server), o NumPy e o SciPy devem estar instalados no servidor. Entre em contato com seu administrador de sistema para obter assistência.

### **Usuários do Windows e Mac**

Para Windows e Mac, o NumPy e SciPy devem ser instalados para uma versão separada do Python 3,10 da versão que está instalada com o IBM SPSS Statistics. Se você não tiver uma versão separada do Python 3,10, será possível fazer download dele a partir do <http://www.python.org>. Em seguida, instale o NumPy e o SciPy para o Python versão 3,10. Os instaladores estão disponíveis a partir do <http://www.scipy.org/Download>.

Para permitir o uso do NumPy e SciPy, deve-se configurar sua localização do Python para a versão do Python 3,10 na qual você instalou o NumPy e o SciPy. O local de Python é configurado a partir da guia Locais do arquivo no diálogo Opções (Editar > Opções).

### **Usuários do Linux**

Sugerimos que você mesmo faça download da origem e construa o NumPy e o SciPy. A origem está disponível a partir do <http://www.scipy.org/Download>. É possível instalar o NumPy e o SciPy para a versão Python 3,10 que é instalada com o IBM SPSS Statistics. Ele está no diretório Python sob o local onde IBM SPSS Statistics está instalado.

Se optar por instalar o NumPy e o SciPy em uma versão do Python 3,10 diferente da versão que está instalada com o IBM SPSS Statistics, então deve-se configurar sua localização do Python para apontar para essa versão. O local do Python é configurado na guia Locais de arquivo da caixa de diálogo Opções (**Editar > Opções**).

### **Usuários do Windows e Unix**

O NumPy e o SciPy devem ser instalados, no servidor, para uma versão separada do Python 3,10 da versão que está instalada com o IBM SPSS Statistics. Se não houver uma versão separada do Python 3,10 no servidor, então ela poderá ser transferida por download a partir do <http://www.python.org>. O NumPy e o SciPy para Python 3,10 estão disponíveis a partir do <http://www.scipy.org/Download>. Para possibilitar o uso do NumPy e SciPy, a localização do Python para o servidor deve ser configurada para a versão do Python 3,10 em que o NumPy e SciPy estão instalados. A localização do Python é configurada a partir do IBM SPSS Statistics Administration Console.



## Modelo

**Especificar Efeitos do Modelo.** Um modelo de efeitos principais contém todos os efeitos principais de fatores e covariáveis. Selecione **Customizado** para especificar interações. Deve-se indicar todos os termos a serem incluídos no modelo.

**Fatores e covariáveis.** Os fatores e covariáveis são listados.

**Modelo.** O modelo depende da natureza de seus dados. Depois de selecionar **Customizado**, é possível selecionar os principais efeitos e interações que são de interesse em sua análise.

Desenvolver termos

Para os fatores e covariáveis selecionados:

**Interação.** Cria o termo de interação de mais alto nível para todas as variáveis selecionadas. Este é o padrão.

**Efeitos principais.** Cria um termo dos principais efeitos para cada variável selecionada.

**Todas 2 fatores.** Cria todas as interações bidirecionais possíveis das variáveis selecionadas.

**Todos os 3 fatores.** Cria todas as interações de três vias possíveis das variáveis selecionadas.

**Todas 4 fatores.** Cria todas as possíveis interações de quatro vias das variáveis selecionadas.

**Todas 5 fatores.** Cria todas as possíveis interações de cinco vias das variáveis selecionadas.

## Opções

A guia Opções permite que o usuário salve e plote estimativas de modelo para casos individuais, fatores latentes e preditores.

Para cada tipo de dados, especifique o nome de um conjunto de dados. Os nomes dos conjuntos de dados devem ser exclusivos. Se você especificar o nome de um conjunto de dados existente, seu conteúdo será substituído; caso contrário, um novo conjunto de dados é criado.

- **Salvar estimativas para casos individuais.** Salva as seguintes estimativas de modelo caso a caso: valores previstos, resíduos, distância até o modelo de fator latente e pontuações de fator latente. Ele também plota pontuações de fatores latentes.
- **Salvar estimativas para fatores latentes.** Salva cargas de fatores latentes e pesos de fatores latentes. Ele também plota os pesos dos fatores latentes.
- **Salvar estimativas para variáveis independentes.** Salva estimativas de parâmetros de regressão e importância variável para projeção (VIP). Também traça o VIP por fator latente.

## Análise do vizinho mais próximo

---

Análise do Vizinho mais Próximo é um método de classificação de casos com base na sua similaridade com outros casos. Em aprendizado por máquina, ela foi desenvolvida como uma maneira de reconhecer padrões de dados sem requerer uma correspondência exata com nenhum dos padrões ou casos armazenados. Casos semelhantes ficam próximos uns dos outros e os casos diferentes ficam distantes uns dos outros. Portanto, a distância entre dois casos é uma medida de sua dissimilaridade.

Casos próximos são chamados de "vizinhos". Quando um novo caso (validação) é apresentado, sua distância de cada um dos casos no modelo é calculada. As classificações dos casos mais similares – os vizinhos mais próximos – são verificadas e o novo caso é colocado na categoria que contiver o maior número de vizinhos mais próximos.

É possível especificar o número de vizinhos mais próximos a serem examinados; este valor é chamado de  $k$ .

A análise do vizinho mais próximo também pode ser utilizada para calcular valores para uma variável resposta contínua. Nesta situação, a média ou mediana do valor dos vizinhos mais próximos é utilizada para obter o valor predito para o novo caso.

Considerações sobre dados de análise do vizinho mais próximo












**Resposta e recursos.** O destino e os recursos podem ser:

- *Nominal.* Uma variável pode ser tratada como nominal quando seus valores representam categorias sem uma classificação intrínseca (por exemplo, o departamento da empresa na qual um funcionário trabalha) Exemplos de variáveis nominais incluem região, código de endereçamento postal e filiação religiosa.
- *Ordinal.* Uma variável pode ser tratada como ordinal quando seus valores representam categorias com alguma classificação intrínseca (por exemplo, níveis de satisfação de serviço de altamente insatisfeito para altamente satisfeito. Exemplos de variáveis ordinais incluem escores de atitude que representam o grau de satisfação ou de confiança e os escores de classificação de preferência.
- *scale.* Uma variável pode ser tratada como escala (contínua) quando seus valores representam categorias ordenadas com uma métrica significativa, para que as comparações de distância entre valores sejam apropriadas. Exemplos de variáveis de escala incluem idade em anos e rendimento em milhares de dólares.

Variáveis nominais e ordinais são tratadas de forma equivalente pela Análise do Vizinho Mais Próximo. O procedimento pressupõe que o nível de medição apropriado foi atribuído a cada variável; no entanto, é possível mudar temporariamente o nível de medição de uma variável clicando com o botão direito do mouse na variável na lista de variáveis de origem e selecionando um nível de medição no menu pop-up.

Um ícone ao lado de cada variável na lista de variáveis identifica o nível de medição e o tipo de dados:

*Tabela 1. Ícones do nível de medição*

	<b>Numérico</b>	<b>Sequência de caracteres</b>	<b>Data</b>	<b>Hora</b>
Escala (Contínua)		n/d		
Ordinal				
Nominal				

**Codificação de variável categórica.** O procedimento recodifica temporariamente preditores categóricos e variáveis dependentes utilizando a codificação um-de-c durante o procedimento. Se houver *c* categorias de uma variável, então a variável será armazenada como vetores *c*, com a primeira categoria denotada (1,0,...,0), a próxima categoria (0,1,0,...,0), ... e a categoria final (0,0,...,0,1).

Este esquema de codificação aumenta a dimensionalidade do espaço de características. Em particular, o número total de dimensões é o número de preditores de escala mais o número de categorias em todos os preditores categóricos. Como resultado, esse esquema de codificação pode levar a um treinamento mais lento. Se o treinamento dos vizinhos mais próximos estiver ocorrendo muito lentamente, você pode tentar reduzir o número de categorias em seus preditores categóricos combinando categorias semelhantes ou descartando casos que tenham categorias extremamente raras antes de executar o procedimento.

Toda a codificação um-de-c baseia-se nos dados de treinamento, mesmo se uma amostra de validação estiver definida (consulte “Partições” na página 172). Assim, se a amostra de validação contiver casos com categorias de predição que não estão presentes nos dados de treinamento, esses casos não serão pontuados. Se a amostra de validação contiver casos com categorias de variáveis dependentes que não estão presentes nos dados de treinamento, esses casos serão pontuados.

**Nova escala.** Os recursos de escala são normalizados por padrão. Todo o reajuste de escala é executado com base nos dados de treinamento, mesmo se uma amostra de validação estiver definida (consulte “Partições” na página 172). Se você especificar uma variável para definir partições, é importante que os recursos tenham distribuições semelhantes nas amostras de treinamento e validação. Utilize, por exemplo, o procedimento [Explorar](#) para examinar as distribuições entre as partições.

**Ponderações de frequência.** As ponderações de frequência são ignoradas por este procedimento.

**Replicando resultados.** O procedimento usa a geração de números aleatórios durante a atribuição aleatória de partições e dobras de validação cruzada. Se desejar replicar seus resultados de modo exato, além de utilizar as mesmas configurações do procedimento, configure um valor semente para o Mersenne Twister (consulte “Partições ” na página 172), ou utilize variáveis para definir partições e dobras de validação cruzada.

Para obter uma análise do vizinho mais próximo

No menu, escolha:

**Analisar > Classificar > Vizinho Mais Próximo...**

1. Especifique um ou mais recursos, que podem ser considerados variáveis independentes ou preditores se houver um destino.

**Destino (opcional).** Se nenhum destino (variável ou resposta) for especificado, então o procedimento localizará somente os  $k$  vizinhos mais próximos, e nenhuma classificação ou predição será feita.

**Normalizar recursos de escala.** As variáveis normalizadas possuem o mesmo intervalo de valores, que pode melhorar o desempenho do algoritmo de estimação. A normalização ajustada,  $[2^*(x-\min)/(\max-\min)]-1$ , é usada. Os valores normalizados ajustados ficam entre -1 e 1.

**Identificador de caso focal (opcional).** Isso permite que você marque casos de interesse particular. Por exemplo, um pesquisador deseja determinar se os resultados dos testes de um distrito escolar – o caso focal – são comparáveis aos de distritos escolares semelhantes. Ele usa a análise do vizinho mais próximo para localizar os distritos escolares mais semelhantes em relação a um determinado conjunto de características. Em seguida, ele compara os resultados dos testes do distrito escolar focal com os dos vizinhos mais próximos.

Casos focais também podem ser usados em estudos clínicos para selecionar casos de controle semelhantes aos casos clínicos. Os casos focais são exibidos nos  $k$  vizinhos mais próximos e na tabela de distância, nos gráficos de espaço de variáveis, no gráfico de peers e no mapa de quadrante. As informações sobre casos focais são salvas nos arquivos especificados na guia Saída.

Casos com um valor positivo na variável especificada são tratados como casos focais. É inválido especificar uma variável sem valores positivos.

**Rótulo case (opcional).** Os casos são rotulados usando esses valores no gráfico de espaço da variável, gráfico de peers e mapa de quadrantes.

Campos com nível de medição desconhecido

O alerta de Nível de Medição é exibido quando o nível de medição para uma ou mais variáveis (campos) no conjunto de dados é desconhecido. Como o nível de medição afeta o cálculo de resultados para este procedimento, todas as variáveis devem ter um nível de medição definido.

**Dados de varredura.** Lê os dados no conjunto de dados ativo e designa o nível de medição padrão para quaisquer campos com um nível de medição desconhecido atualmente. Se o conjunto de dados for grande, isso poderá demorar algum tempo.

**Designar Manualmente.** Abre um diálogo que lista todos os campos com um nível de medição desconhecido. É possível utilizar este diálogo para designar o nível de medição para esses campos. Também é possível designar o nível de medição na Visualização de Variável do Editor de Dados.

Como o nível de medição é importante para este procedimento, não é possível acessar o diálogo para executar este procedimento até que todos os campos possuam um nível de medição definido.

## Vizinhos

**Número de Vizinhos Mais Próximos (k).** Especifique o número de vizinhos mais próximos. Observe que utilizar um número maior de vizinhos não resultará necessariamente em um modelo mais preciso.

Se um destino for especificado na guia Variáveis, você pode, alternativamente, especificar um intervalo de valores e permitir que o procedimento escolha o "melhor" número de vizinhos dentro desse intervalo.

O método para determinar o número de vizinhos mais próximos depende se a seleção de recursos é solicitada na guia Recursos.

- Se a seleção de variável estiver em vigor, então a seleção de variável será executada para cada valor de  $k$  no intervalo solicitado, e o  $k$ , junto do conjunto de recursos acompanhante, com a menor taxa de erro (ou a menor soma dos quadrados dos erros se a resposta for escalar) será selecionado.
- Se a seleção de variável não estiver em vigor, então a validação cruzada da dobra  $V$  será utilizada para selecionar o "melhor" número de vizinhos. Consulte a guia Partição para controlar a atribuição de dobras.

**Cálculo de Distância.** Esta é a métrica utilizada para especificar a métrica de distância utilizada para medir a similaridade de casos.

- **Métrica euclidiana.** A distância entre dois casos,  $x$  e  $y$ , é a raiz quadrada da soma, sobre todas as dimensões, das diferenças quadradas entre os valores dos casos.
- **Métrica de City Block.** A distância entre dois casos é a soma, em todas as dimensões, das diferenças absolutas entre os valores dos casos. Essa métrica também é chamada de distância de Manhattan.

Opcionalmente, se um destino for especificado na guia Variáveis, será possível optar por ponderar os recursos por sua importância normalizada ao calcular as distâncias. A importância do recurso para um preditor é calculada pela razão da taxa de erro ou erro de soma de quadrados do modelo com o preditor removido do modelo para a taxa de erro ou erro de soma de quadrados para o modelo completo. A importância normalizada é calculada reponderando os valores de importância do recurso para que somam 1.

**Predições para Variável Resposta Escalar.** Se um alvo de escala for especificado na guia Variáveis, isso especifica se o valor previsto é calculado com base na média ou no valor mediano dos vizinhos mais próximos.

## Recursos

A guia Recursos permite solicitar e especificar opções para seleção de recursos quando um destino é especificado na guia Variáveis. Por padrão, todos os recursos são considerados para seleção de recursos, mas você pode, opcionalmente, selecionar um subconjunto de recursos para forçar o modelo.

**Critério de Parada.** Em cada etapa, o recurso cuja adição ao modelo resulta no menor erro (calculado como a taxa de erro para um alvo categórico e o erro de soma dos quadrados para um alvo de escala) é considerado para inclusão no conjunto de modelos. A seleção Forward continua até que a condição especificada seja atendida.

- **Número de recursos especificados.** O algoritmo inclui um número fixo de recursos além daqueles forçados no modelo. Especifique um número inteiro positivo. Diminuir os valores do número a ser selecionado cria um modelo mais parcimonioso, com o risco de perder características importantes. Aumentar os valores do número a ser selecionado capturará todos os recursos importantes, com o risco de eventualmente adicionar recursos que realmente aumentam o erro do modelo.
- **Mudança mínima na razão de erro absoluto.** O algoritmo para quando a mudança na razão de erro absoluta indica que o modelo não pode ser melhorado adicionando mais recursos. Especifique um número positivo. Diminuir os valores da mudança mínima tenderá a incluir mais recursos, sob o risco de incluir recursos que não agregam muito valor ao modelo. Aumentar o valor da mudança mínima tenderá a excluir mais recursos, correndo o risco de perder recursos que são importantes para o modelo. O valor "ideal" da mudança mínima dependerá de seus dados e aplicativo. Consulte o log de erros de seleção de recursos na saída para ajudá-lo a avaliar quais recursos são mais importantes. Consulte o tópico "[Log de erro de seleção de variável](#)" na página 177 para obter mais informações

## Partições

A guia Partições permite dividir o conjunto de dados em conjuntos de treinamento e validação e, quando aplicável, atribuir casos em pastas de validação cruzada

**Partições de Treinamento e de Validação.** Este grupo especifica o método de particionamento do conjunto de dados ativo em amostras de treinamento e validação. A **amostra de treinamento**

compreende os registros de dados utilizados para treinar o modelo de vizinho mais próximo, e alguma porcentagem de casos no conjunto de dados deve ser designada para a amostra de treinamento para obter um modelo. A **amostra de validação** é um conjunto independente de registros de dados utilizados para avaliar o modelo final; o erro para a amostra de validação fornece uma estimativa "honesta" da capacidade preditiva do modelo porque os casos de validação não foram utilizados para construir o modelo.

- **Designar aleatoriamente casos às partições.** Especifique a porcentagem de casos a serem atribuídos à amostra de treinamento. O restante é atribuído à amostra de validação.
- **Usar variável para designar casos.** Especifique uma variável numérica que atribui cada caso no conjunto de dados ativo à amostra de treinamento ou validação. Casos com um valor positivo na variável são atribuídos à amostra de treinamento, casos com valor 0 ou negativo, à amostra de validação. Casos com um valor omissivo do sistema são excluídos da análise. Quaisquer valores ausentes de usuário para a variável de partição são sempre tratados como válidos.

**Dobras de Validação Cruzada.** A validação cruzada de dobra  $V$  é utilizada para determinar o "melhor" número de vizinhos. Ela não está disponível junto com a seleção de variável por motivos de desempenho.

A validação cruzada divide a amostra em um número de subamostras ou dobras. Os modelos do vizinho mais próximo são então gerados, excluindo os dados de cada subamostra por vez. O primeiro modelo é baseado em todos os casos, exceto aqueles na primeira dobra de amostra, o segundo modelo é baseado em todos os casos, exceto aqueles na segunda dobra de amostra, e assim por diante. Para cada modelo, o erro é estimado aplicando-se o modelo à subamostra excluída na sua geração. O "melhor" número de vizinhos mais próximos é aquele que produz o menor erro nas dobras.

- **Designar aleatoriamente casos às dobras.** Especifique o número de dobras que devem ser usadas para validação cruzada. O procedimento designa aleatoriamente casos às dobras, numerados de 1 a  $V$ , o número de dobras.
- **Usar variável para designar casos.** Especifique uma variável numérica que atribui cada caso no conjunto de dados ativo a uma dobra. A variável deve ser numérica e assumir valores de 1 a  $V$ . Se algum valor neste intervalo estiver ausente e em qualquer divisão se os arquivos divididos estiverem em vigor, isso causará um erro. para obter mais informações.

**Configurar semente para Mersenne Twister.** Configurar uma semente permite replicar análises. Usar este controle é semelhante a definir o Mersenne Twister como o gerador ativo e especificar um ponto de partida fixo na caixa de diálogo Geradores de números aleatórios, com a importante diferença de que definir a semente nesta caixa de diálogo preservará o estado atual do gerador de números aleatórios e restaurará esse estado após a conclusão da análise. para obter mais informações.

## Salvar

**Nomes de Variáveis Salvas.** A geração automática de nomes assegura que você mantenha todo o seu trabalho. Os nomes customizados permitem descartar/substituir resultados de execuções anteriores sem primeiro excluir as variáveis salvas no Editor de dados.

Variáveis para salvar

- **Valor predito ou categórico.** Isso salva o valor previsto para um alvo de escala ou a categoria prevista para um alvo categórico.
- **Probabilidade predita.** Isso salva as probabilidades previstas para um alvo categórico. Uma variável separada é salva para cada uma das  $n$  primeiras categorias, em que  $n$  é especificado no controle **Máximo de categorias a serem salvas para variável resposta categórica**.
- **Variáveis de partição de Treinamento/Validação.** Se os casos forem atribuídos aleatoriamente às amostras de treinamento e validação na guia Partições, isso salvará o valor da partição (treinamento ou validação) à qual o caso foi atribuído.
- **Variável de dobra de validação cruzada.** Se os casos forem atribuídos aleatoriamente a dobras de validação cruzada na guia Partições, isso salvará o valor da dobra à qual o caso foi atribuído.

## Saída

Saída do visualizador

- **Sumarização do processamento de caso.** Exibe a tabela de resumo de processamento de casos, que resume o número de casos incluídos e excluídos na análise, no total e por amostras de treinamento e validação.
- **Gráficos e tabelas.** Exibe a saída relacionada ao modelo, incluindo tabelas e gráficos. As tabelas na visualização de modelo incluem  $k$  vizinhos mais próximos e distâncias para casos focais, classificação de variáveis resposta categórica e uma sumarização de erros. A saída gráfica na visualização do modelo inclui um log de erros de seleção, gráfico de importância de recurso, gráfico de espaço de recurso, gráfico de pares e mapa de quadrante. Consulte o tópico “[Visualização do Modelo](#)” na página 174 para obter mais informações

Arquivos

- **Exportar modelo em XML.** É possível usar esse arquivo de modelo para aplicar as informações de modelo a outros arquivos de dados para propósitos de escoragem. para obter mais informações. Esta opção não está disponível se os arquivos divididos tiverem sido definidos.
- **Exportar distâncias entre casos focais e  $k$  vizinhos mais próximos.** Para cada caso focal, uma variável separada é criada para cada um dos  $k$  vizinhos mais próximos do caso focal (a partir da amostra de treinamento) e para as  $k$  distâncias mais próximas correspondentes.

## Opções

**Valores Omissos de Usuário.** As variáveis categóricas devem ter valores válidos para que um caso seja incluído na análise. Esses controles permitem que você decida se os valores ausentes do usuário são tratados como válidos entre as variáveis categóricas.

Valores ausentes do sistema e valores omissos para variáveis de escala são sempre tratados como inválidos.

## Visualização do Modelo

Ao selecionar **Gráficos e tabelas** na guia Saída, o procedimento cria um objeto Modelo de Vizinho Mais Próximo no Visualizador. Ao ativar (dar um clique duplo) neste objeto, você obtém uma visualização interativa do modelo. A visualização do modelo possui uma janela de 2 painéis:

- O primeiro painel exibe uma visão geral do modelo chamada de visualização principal.
- O segundo painel exibe um dos dois tipos de visualizações:

Uma visualização de modelo auxiliar mostra mais informações sobre o modelo, mas não é focada no modelo em si.

Uma visualização vinculada é uma visualização que mostra detalhes sobre um recurso do modelo quando o usuário detalha parte da visualização principal.

Por padrão, o primeiro painel mostra o espaço de recurso e o segundo painel mostra o gráfico de importância de variável. Se o gráfico de importância de variável não estiver disponível, ou seja, quando **Ponderar recursos por importância** não foi selecionada na guia Recursos, a primeira visualização disponível no menu suspenso Visualizar será mostrada.

Quando uma visualização não possui informações disponíveis, o texto do item na lista suspensa Exibir é desativado.

## Espaço da Variável

O gráfico de espaço de feição é um gráfico interativo do espaço de feição (ou um subespaço, se houver mais de 3 feições). Cada eixo representa um recurso no modelo e a localização dos pontos no gráfico mostra os valores desses recursos para casos nas partições de treinamento e validação.

**Chaves.** Além dos valores dos recursos, os pontos no gráfico transmitem outras informações.

- A forma indica a partição à qual um ponto pertence, seja Treinamento ou Validação.
- A cor/sombreamento de um ponto indica o valor do alvo para aquele caso; com valores de cores distintos iguais às categorias de um alvo categórico e tons que indicam o intervalo de valores de um alvo contínuo. O valor indicado para a partição de treinamento é o valor observado; para a partição de validação, é o valor predito. Se nenhum destino for especificado, essa chave não será mostrada.
- Estruturas de tópicos mais pesadas indicam que um caso é focal. Casos focais são mostrados vinculados aos seus  $k$  vizinhos mais próximos.

**Controles e Interatividade.** Vários controles no gráfico permitem que você explore o espaço de recurso.

- Você pode escolher qual subconjunto de recursos mostrar no gráfico e alterar quais recursos são representados nas dimensões.
- “Casos focais” são simplesmente pontos selecionados no gráfico de espaço da variável. Se você especificou uma variável de caso focal, os pontos que representam os casos focais serão inicialmente selecionados. No entanto, qualquer ponto pode se tornar temporariamente um caso focal se você o selecionar. Os controles “usuais” para seleção de pontos se aplicam; clicar em um ponto seleciona esse ponto e desmarca todos os outros; Clicar com a tecla Control pressionada em um ponto o adiciona ao conjunto de pontos selecionados. As visualizações vinculadas, como o Gráfico de peers, serão atualizadas automaticamente com base nos casos selecionados no Espaço de recursos.
- É possível alterar o número de vizinhos mais próximos ( $k$ ) para exibir para casos focais.
- Passar o mouse sobre um ponto no gráfico exibe uma dica de ferramenta com o valor do rótulo do caso ou número do caso se os rótulos do caso não estiverem definidos e os valores alvo observados e previstos.
- Um botão “Reconfigurar” permite que você retorne o espaço de recurso ao seu estado original.

### ***Incluindo e removendo arquivos/variáveis***

É possível incluir novos campos/variáveis ao espaço de feição ou remover os que estão sendo exibidos no momento.

Paleta de variáveis

A paleta de variáveis deve ser exibida antes que seja possível incluir e remover variáveis. Para exibir a paleta de variáveis, o Visualizador de modelo deve estar em modo de edição e um caso deve ser selecionado no espaço do recurso.

1. Para colocar o Visualizador de modelo em modo de edição, a partir dos menus, escolha:

**Visualização > Modo de edição**

2. Uma vez no modo de edição, clique em qualquer caso no espaço de recurso.
3. Para exibir a paleta de variáveis, a partir dos menus, escolha:

**Visualização > Paletas > Variáveis**

A paleta de variáveis lista todas as variáveis no espaço do recurso. O ícone ao lado do nome da variável indica o nível de medição da variável.

4. Para alterar temporariamente o nível de medição de uma variável, clique com o botão direito na variável na paleta de variáveis e escolha uma opção.

Zonas de variáveis

As variáveis são incluídas em "zonas" no espaço do recurso. Para exibir as zonas, inicie arrastando uma variável da paleta de variáveis ou selecione **Mostrar zonas**.

O espaço do recurso possui zonas para os eixos  $x$ ,  $y$  e  $z$ .

Movendo variáveis nas zonas

Aqui estão algumas regras gerais e dicas para mover variáveis para as zonas:

- Para mover uma variável para uma zona, clique e arraste a variável da paleta de variáveis e solte-a na zona. Se você escolher **Mostrar zonas**, será possível também clicar com o botão direito do mouse em uma zona e selecione uma variável que você deseja incluir na zona.
- Se você arrastar uma variável da paleta Variáveis para uma zona já ocupada por outra variável, a variável antiga será substituída pela nova.
- Se você arrastar uma variável de uma zona para uma zona já ocupada por outra variável, as variáveis trocam de posição.
- Clicar no X em uma zona remove a variável dessa zona.
- Se houver vários elementos gráficos na visualização, cada elemento gráfico pode ter suas próprias zonas de variáveis associadas. Primeiro, selecione o elemento gráfico.

## Importância da Variável

Normalmente, você desejará concentrar seus esforços de modelagem nas variáveis que mais importam e considerar descartar ou ignorar aquelas que menos importam. O gráfico de importância variável ajuda você a fazer isso, indicando a importância relativa de cada variável na estimativa do modelo. Como os valores são relativos, a soma dos valores de todas as variáveis na tela é 1,0. A importância da variável não está relacionada à precisão do modelo. Relaciona-se apenas com a importância de cada variável em fazer uma previsão, não se a previsão é precisa ou não.

## Peers

Este gráfico exibe os casos focais e seus  $k$  vizinhos mais próximos em cada recurso e na resposta. Ele estará disponível se um caso focal for selecionado no Espaço de Recursos.

**Vinculando comportamento.** O gráfico Peers é vinculado ao Espaço da Variável de duas maneiras.

- Casos selecionados (focais) no Espaço da Variável são exibidos no gráfico de Peers, junto de seus  $k$  vizinhos mais próximos.
- O valor de  $k$  selecionado no Espaço Recurso é usado no gráfico de Peers.

## Distâncias de Vizinho Mais Próximo

Esta tabela exibe os  $k$  vizinhos mais próximos e as distâncias somente para casos focais. Ela estará disponível se um identificador de caso focal for especificado na guia Variáveis, e exibe somente os casos focais identificados por essa variável.

Cada linha da:

- A coluna **Caso Focal** contém o valor da variável de rótulo case para o caso focal; se rótulos case não forem definidos, essa coluna conterá o número do caso do caso focal.
- A  $i$ -ésima coluna no grupo Vizinhos Mais Próximos contém o valor da variável de rotulagem de caso para o  $i$ -ésimo vizinho mais próximo do caso focal; se rótulos case não forem definidos, esta coluna conterá o número do caso do  $i$ -ésimo vizinho mais próximo do caso focal.
- A  $i$ -ésima coluna no grupo Distâncias Mais Próximas contém a distância do  $i$ -ésimo vizinho mais próximo do caso focal

## Mapa de quadrante

Este gráfico exibe os casos focais e seus  $k$  vizinhos mais próximos em um gráfico de dispersão (ou gráfico de pontos, dependendo do nível de medição da resposta) com a resposta no eixo  $y$  e uma variável de escala no eixo  $x$ , agrupados em painéis por variáveis. Ele estará disponível se houver uma resposta e se um caso focal estiver selecionado no Espaço de Variável.

- As linhas de referência são desenhadas para variáveis contínuas, nas médias da variável na partição de treinamento.



## Log de erro de seleção de variável

Pontos no gráfico exibem o erro (a taxa de erros ou a soma dos quadrados dos erro, dependendo do nível de medição da resposta) no eixo  $y$  para o modelo com o recurso listado no eixo  $x$  (além de todos os recursos à esquerda no eixo  $x$ ). Este gráfico está disponível se houver um destino, e se a seleção de recursos estiver em vigor.

## Log de erro de seleção $k$

Pontos no gráfico exibem o erro (a taxa de erros ou a soma dos quadrados dos erros, dependendo do nível de medição da resposta) no eixo  $y$  para o modelo com o número de vizinhos mais próximos ( $k$ ) no eixo  $x$ . Este gráfico estará disponível se houver uma resposta e uma seleção  $k$  estiver em vigor.

## Log de Erro de $k$ e Seleção de variável

Estes são gráficos de seleção de recursos (consulte “[Log de erro de seleção de variável](#)” na página 177), em painéis por  $k$ . Este gráfico está disponível se houver um destino, e  $k$  e a seleção de recursos estiverem em vigor.

## Tabela de Classificação

Esta tabela exibe a classificação cruzada de valores observados versus valores preditos da resposta, por partição. Está disponível se houver um destino e for categórico.

- A linha (**Omisso**) na partição Validação contém casos de validação com valores omissos na resposta. Esses casos contribuem para os valores de Amostra de Holdout: Percentual Geral, mas não para os valores de Percentual Correto.

## Sumarização de Erro

Esta tabela está disponível se houver uma variável de destino. Exibe o erro associado ao modelo; soma de quadrados para uma destino contínuo e a taxa de erro (100% – porcentagem geral correta) para uma destino categórico.

## Análise discriminante

A análise discriminante constrói um modelo preditivo para associação ao grupo. O modelo é composto por uma função discriminante (ou, para mais de dois grupos, um conjunto de funções discriminantes) baseada em combinações lineares das variáveis preditoras que proporcionam a melhor discriminação entre os grupos. As funções são geradas a partir de uma amostra de casos para os quais a associação ao grupo é conhecida; as funções podem então ser aplicadas a novos casos que possuem medidas para as variáveis preditoras, mas possuem associação de grupo desconhecida.

*Nota:* A variável de agrupamento pode ter mais de dois valores. Os códigos para a variável de agrupamento devem ser inteiros, no entanto, e você precisa especificar seus valores mínimo e máximo. Casos com valores fora desses limites são excluídos da análise.

**Exemplo.** Em média, as pessoas nos países da zona temperada consomem mais calorias por dia do que as pessoas nos trópicos, e uma proporção maior das pessoas nas zonas temperadas são moradores das cidades. Um pesquisador deseja combinar essas informações em uma função para determinar quão bem um indivíduo pode discriminar entre os dois grupos de países. O pesquisador acha que o tamanho da população e as informações econômicas também podem ser importantes. A análise discriminante permite estimar os coeficientes da função discriminante linear, que se parece com o lado direito de uma equação de regressão linear múltipla. Ou seja, utilizando os coeficientes  $a$ ,  $b$ ,  $c$  e  $d$ , a função é:

$$D = a * climate + b * urban + c * population + d * gross domestic product per capita$$

Se essas variáveis forem úteis para discriminar entre as duas zonas climáticas, os valores de  $D$  irão diferir para países de zonas temperadas e tropicais. Se você usar um método de seleção de variável passo a passo, poderá descobrir que não precisa incluir todas as quatro variáveis na função.

**Estatísticas.** Para cada variável: médias, desvios padrão, ANOVA univariada. Para cada análise: *M* de Box, matriz de correlações dentro de grupos, matriz de covariâncias dentro de grupos, matriz de covariâncias de grupos separados, matriz de covariâncias totais. Para cada função discriminante canônica: autovalor, porcentagem de variância, correlação canônica, lambda de Wilks, qui-quadrado. Para cada etapa: probabilidades anteriores, coeficientes de função de Fisher, coeficientes de função não padronizados, lambda de Wilks para cada função canônica.

Considerações de Dados de Análise Discriminante

**Dados.** A variável de agrupamento deve ter um número limitado de categorias distintas, codificadas como números inteiros. As variáveis independentes que são nominais devem ser recodificadas como variáveis dummy ou de contraste.

**Suposições.** Os casos devem ser independentes. As variáveis preditoras devem ter uma distribuição normal multivariada e as matrizes de variância-covariância dentro do grupo devem ser iguais entre os grupos. A associação de grupo é considerada mutuamente exclusiva (ou seja, nenhum caso pertence a mais de um grupo) e coletivamente exaustiva (ou seja, todos os casos são membros de um grupo). O procedimento é mais eficaz quando a associação ao grupo é uma variável verdadeiramente categórica; se a associação ao grupo for baseada em valores de uma variável contínua (por exemplo, QI alto versus QI baixo), considere usar a regressão linear para aproveitar as informações mais ricas oferecidas pela própria variável contínua.

Para obter uma análise discriminante

1. A partir dos menus, escolha:

**Analisar > Classificar > Discriminante...**

2. Selecione uma variável de agrupamento com valor de número inteiro e clique em **Definir Intervalo** para especificar as categorias de interesse.

3. Selecione as variáveis independentes ou preditoras. (Se a sua variável de agrupamento não tiver valores inteiros, a Recodificação Automática no menu Transformar criará uma variável que tenha.)

4. Selecione o método para inserir as variáveis independentes.

- **Inserir independentes juntas.** Insere simultaneamente todas as variáveis independentes que atendem aos critérios de tolerância.
- **Utilizar o método stepwise.** Usa análise de stepwise para controlar a entrada e a remoção de variáveis.

5. Opcionalmente, selecione casos com uma variável de seleção.

## Análise discriminante Definir intervalo

Especifique o valor mínimo e máximo da variável de agrupamento para a análise. Casos com valores fora dessa faixa não são utilizados na análise discriminante, mas são classificados em um dos grupos existentes com base nos resultados da análise. Os valores mínimo e máximo devem ser números inteiros.

## Casos Selecionados de Análise Discriminante

Para selecionar casos para a sua análise:

1. Na caixa de diálogo Análise discriminante, escolha uma variável de seleção.

2. Clique em **Valor** para inserir um número inteiro como o valor de seleção.

Apenas casos com o valor especificado para a variável de seleção são utilizados para derivar as funções discriminantes. Os resultados das estatísticas e de classificação são gerados para casos selecionados e não selecionados. Esse processo fornece um mecanismo para classificar novos casos com base em dados existentes anteriormente ou para particionar seus dados em subconjuntos de treinamento e teste para realizar a validação no modelo gerado.

## Estatísticas de análise discriminante

**Descritivos.** As opções disponíveis são médias (incluindo desvios padrão), ANOVAs univariadas e teste *M* de Box.

- *Médias.* Exibe as médias totais e de grupo, bem como os desvios padrão para as variáveis independentes.
- *ANOVAs Univariadas.* Executa um teste de análise de variância unidirecional de igualdade de médias de grupo para cada variável independente.
- *M de Box.* Um teste para a igualdade das matrizes de covariância do grupo. Para amostras suficientemente grandes, um valor *p* não significativo quer dizer que não há evidência suficiente para as matrizes diferirem. O teste é sensível a partidas da normalidade multivariada.

**Coefficientes de Função.** As opções disponíveis são coeficientes de classificação de Fisher e coeficientes não padronizados.

- *de Fisher.* Exibe os coeficientes de função de classificação do Fisher que podem ser usados diretamente para classificação. Um conjunto separado de coeficientes de função de classificação é obtido para cada grupo, e um caso é designado ao grupo para o qual ele possui o maior escore discriminante (valor da função de classificação).
- *Não padronizado.* Exibe os coeficientes de função discriminante não padronizadas.

**Matrizes.** As matrizes de coeficientes disponíveis para variáveis independentes são matriz de correlações dentro de grupos, matriz de covariâncias dentro de grupos, matriz de covariâncias de grupos separados e matriz de covariâncias totais.

- *Correlação intragrupos.* Exibe uma matriz de correlações dentro de grupos em conjunto que é obtida pela média das matrizes de covariâncias separadas de todos os grupos antes de calcular as correlações.
- *Covariância intragrupos.* Exibe uma matriz de covariância agrupada dentro de grupos, que pode diferir da matriz de covariâncias totais. A matriz é obtida pela média das matrizes de covariâncias separadas para todos os grupos.
- *Covariância de grupos separados.* Exibe matrizes de covariância separadas para cada grupo.
- *Covariância total.* Exibe uma matriz de covariâncias a partir de todos os casos como se fossem de uma única amostra.

## Método de análise discriminante stepwise

**Método.** Selecione a estatística a ser utilizada para inserir ou remover novas variáveis. As alternativas disponíveis são lambda de Wilks, variância inexplicada, distância de Mahalanobis, menor razão *F* e *V* de Rao. Com *V* de Rao, é possível especificar o aumento mínimo em *V* para uma variável entrar.

- *Lambda de Wilks.* Um método de seleção de variáveis para análise discriminante stepwise que escolhe as variáveis para entrada na equação com base em quanto elas reduzem o lambda de Wilks. Em cada passo, a variável que minimiza o lambda geral de Wilks é inserida.
- *Variância inexplicada.* Em cada passo, a variável que minimiza a soma da variação não explicada entre os grupos é inserida.
- *Distância Mahalanobis.* Uma medida de quanto os valores de um caso nas variáveis independentes diferem da média de todos os casos. Uma distância de Mahalanobis grande identifica um caso como tendo valores extremos em uma ou mais variáveis independentes.
- *Menor proporção F.* Um método de seleção de variáveis na análise stepwise com base na maximização de uma razão *F* calculada a partir da distância de Mahalanobis entre os grupos.
- *V de Rao.* Uma medida das diferenças entre as médias do grupo. Também chamada de rastreamento de Lawley-Hotelling. Em cada passo, a variável que maximiza o aumento no *V* de Rao é inserida. Após selecionar esta opção, insira o valor mínimo que uma variável deve ter para inserir na análise.

**Crítérios.** As alternativas disponíveis são o **valor de F de uso** e a **probabilidade de uso de F**. Insira valores para inserir e remover variáveis.

- *Usar valor de F.* Uma variável será inserida no modelo se seu valor F for maior que o valor de Entrada e será removida se o valor F for menor que o valor de Remoção. A Entrada deve ser maior que Remoção, e ambos os valores devem ser positivos. Para inserir mais variáveis no modelo, diminua o valor de Entrada. Para remover mais variáveis do modelo, aumente o valor de Remoção.
- *Usar probabilidade de F.* Uma variável será inserida no modelo se o nível de significância de seu valor F for menor que o valor de entrada e será removida se o nível de significância for maior que o valor de Remoção. A Entrada deve ser menor que Remoção, e ambos os valores devem ser positivos. Para inserir mais variáveis no modelo, aumente o valor de Entrada. Para remover mais variáveis do modelo, diminua o valor de Remoção.

**Exibição. Resumo das etapas** exibe estatísticas para todas as variáveis após cada etapa; **F para distâncias pareadas** exibe uma matriz de razões *F* pareadas para cada par de grupos.

## Classificação da Análise Discriminante

**Probabilidades Anteriores.** Esta opção determina se os coeficientes de classificação são ajustados para um conhecimento a priori de associação ao grupo.

- **Todos os grupos iguais.** Probabilidades anteriores iguais são assumidas para todos os grupos; isso não tem efeito sobre os coeficientes.
- **Calcular a partir de tamanhos de grupo.** Os tamanhos de grupo observados em sua amostra determinam as probabilidades anteriores de associação ao grupo. Por exemplo, se 50% das observações incluídas na análise se enquadram no primeiro grupo, 25% no segundo e 25% no terceiro, os coeficientes de classificação são ajustados para aumentar a probabilidade de pertencer ao primeiro grupo em relação ao outros dois.

**Exibição.** As opções de exibição disponíveis são resultados caso a caso, tabela de resumo e classificação de exclusão.

- *Resultados do casewise.* Os códigos para o grupo real, o grupo predito, as probabilidades posteriores e os escores discriminantes são exibidos para cada caso.
- *Tabela de resumo.* O número de casos corretamente e incorretamente designados a cada um dos grupos com base na análise discriminante. Às vezes é chamado de "Matriz de Confusão".
- *Classificação com exclusão de um item.* Cada caso na análise é classificado pelas funções derivadas de todos os casos diferentes daquele caso. Também é conhecida como "Método U".

**Substituir valores omissos pela média.** Selecione esta opção para substituir a média de uma variável independente por um valor ausente apenas durante a fase de classificação.

**Usar Matriz de Covariâncias.** É possível optar por classificar casos utilizando uma matriz de covariâncias dentro de grupos ou uma matriz de covariâncias separada de grupos.

- *Dentro de Grupos.* A matriz de covariâncias dentro de grupos em conjunto é usada para classificar casos.
- *Grupos separados.* Matrizes de covariâncias de grupos separados são usadas para classificação. Como a classificação é baseada nas funções discriminantes (não com base nas variáveis originais), essa opção nem sempre é equivalente à discriminação quadrática.

**Gráficos.** As opções de plotagem disponíveis são grupos combinados, grupos separados e mapa territorial.

- *Grupos combinados.* Cria um gráfico de dispersão de todos os grupos dos dois primeiros valores da função discriminante. Se houver apenas uma função, um histograma será exibido.
- *Grupos separados.* Cria gráficos de dispersão de grupos separados dos dois primeiros valores da função discriminante. Se houver apenas uma função, histogramas serão exibidos.
- *Mapa territorial.* Um gráfico dos limites usados para classificar os casos em grupos com base em valores da função. Os números correspondem aos grupos nos quais os casos são classificados. A média de cada grupo é indicada por um asterisco dentro de seus limites. O mapa não será exibido se houver apenas uma função discriminante.

## Salvamento de análise discriminante

É possível incluir novas variáveis em seu arquivo de dados ativos. As opções disponíveis são associação prevista ao grupo (uma única variável), pontuações discriminantes (uma variável para cada função discriminante na solução) e probabilidades de associação ao grupo dadas as pontuações discriminantes (uma variável para cada grupo).

Também é possível exportar informações de modelo para o arquivo especificado no formato XML. É possível usar esse arquivo de modelo para aplicar as informações de modelo a outros arquivos de dados para propósitos de escoragem. para obter mais informações.

## Recursos adicionais do comando DISCRIMINANT

O idioma da sintaxe de comando também permite:

- Executar diversas análises discriminante (com um comando) e controlar a ordem na qual as variáveis são inseridas (com o subcomando ANALYSIS).
- Especificar as probabilidades anteriores para classificação (com o subcomando PRIORS).
- Exibir padrões e matrizes de estrutura girados (com o subcomando ROTATE).
- Limitar o número de funções discriminantes extraídas (com o subcomando FUNCTIONS).
- Restringir classificação para os casos que forem selecionados (ou desmarcados) para a análise (com o subcomando SELECT).
- Ler e analisar uma matriz de correlações (com o subcomando MATRIX).
- Gravar uma matriz de correlações para análise posterior (com o subcomando MATRIX).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Análise fatorial

A análise fatorial tenta identificar variáveis subjacentes, ou **fatores**, que explicam o padrão de correlações dentro de um conjunto de variáveis observadas. A análise fatorial é frequentemente usada na redução de dados para identificar um pequeno número de fatores que explicam a maior parte da variância observada em um número muito maior de variáveis manifest. A análise fatorial também pode ser usada para gerar hipóteses sobre mecanismos causais ou para selecionar variáveis para análise posterior (por exemplo, para identificar colinearidade antes de realizar uma análise de regressão linear).

O procedimento de análise fatorial oferece um alto grau de flexibilidade:

- Sete métodos de extração de fator estão disponíveis.
- Cinco métodos de rotação estão disponíveis, incluindo oblímico direto e promax para rotações não ortogonais.
- Estão disponíveis três métodos de cálculo dos escores dos fatores, e os escores podem ser salvos como variáveis para análise posterior.

**Exemplo.** Quais atitudes subjacentes levam as pessoas a responder às perguntas de uma pesquisa de opinião política como o fazem? Ao examinar as correlações entre os itens da pesquisa revela-se que há uma sobreposição significativa entre vários subgrupos de itens – perguntas sobre impostos tendem a se correlacionar, perguntas sobre questões militares se correlacionam, e assim por diante. Com a análise fatorial, é possível investigar o número de fatores subjacentes e, em muitos casos, identificar o que os fatores conceitualmente representam. Além disso, é possível calcular as pontuações dos fatores para cada respondente, que podem ser usadas em análises subsequentes. Por exemplo, você pode construir um modelo de regressão logística para prever o comportamento de votação com base nas pontuações dos fatores.

**Estatísticas.** Para cada variável: número de casos válidos, média e desvio padrão. Para cada análise fatorial: matriz de correlação das variáveis, incluindo níveis de significância, determinante e inversa; matriz de correlação reproduzida, incluindo anti-imagem; solução inicial (comunidades, autovalores e porcentagem de variância explicada); medida de adequação amostral de Kaiser-Meyer-Olkin e teste

de esfericidade de Bartlett; solução não rotacionada, incluindo cargas fatoriais, comunalidades e autovalores; e solução girada, incluindo matriz de padrão rotacionada e matriz de transformação. Para rotações oblíquas: padrão girado e matrizes de estrutura; matriz de coeficiente de pontuação fatorial e matriz de covariância fatorial. Gráficos: gráfico de escarpa de autovalores e gráfico de carregamento dos primeiros dois ou três fatores.

Considerações de Dados de Análise Fatorial

**Dados.** As variáveis devem ser quantitativas no nível de *intervalo* ou *razão*. Dados categóricos (como religião ou país de origem) não são adequados para análise fatorial. Os dados para os quais os coeficientes de correlação de Pearson podem ser calculados de forma sensata devem ser adequados para a análise fatorial.

**Suposições.** Os dados devem ter uma distribuição normal bivariada para cada par de variáveis, e as observações devem ser independentes. O modelo de análise fatorial especifica que as variáveis são determinadas por fatores comuns (os fatores estimados pelo modelo) e fatores únicos (que não se sobrepõem entre as variáveis observadas); as estimativas calculadas são baseadas na suposição de que todos os fatores únicos não são correlacionados entre si e com os fatores comuns.

Para obter uma análise fatorial

1. A partir dos menus, escolha:

**Analisar > Redução de Dimensão > Fator...**

2. Selecione as variáveis para a análise fatorial.

## Selecionar Casos para Análise Fatorial

Para selecionar casos para sua análise:

1. Escolha uma variável de seleção.
2. Clique em **Valor** para inserir um número inteiro como o valor de seleção.

Apenas casos com esse valor para a variável de seleção são usados na análise fatorial.

## Descritivas de Análise Fatorial

**Estatísticas. Descritivos univariados** inclui a média, o desvio-padrão e o número de casos válidos para cada variável. **Solução inicial** exhibe comunalidades iniciais, autovalores e a porcentagem da variância explicada.

**Matriz de Correlações.** As opções disponíveis são coeficientes, níveis de significância, determinante, KMO e teste de esfericidade de Bartlett, inverso, reproduzido e anti-imagem.

- *KMO e Teste de Esfericidade de Bartlett.* A medida Kaiser-Meyer-Olkin de adequação de amostragem testa se as correlações parciais entre variáveis são pequenas. O teste de esfericidade de Bartlett testa se a matriz de correlações é uma matriz de identidade, que indicaria que o modelo de fator é inadequado.
- *Reproduzido.* A matriz de correlações estimada da solução do fator. Os resíduos (diferença entre correlações estimadas e observadas) também são exibidos.
- *Anti-imagem.* A matriz de correlações anti-imagem contém os negativos dos coeficientes de correlação parciais e a matriz de covariâncias anti-imagem contém os negativos das covariâncias parciais. Em um fator de modelo bom, muitos dos elementos fora da diagonal serão pequenos. A medida da adequação de amostragem para uma variável é exibida na diagonal da matriz de correlações anti-imagem.

## Extração de Análise Fatorial

**Método.** Permite especificar o método de extração de fator. Os métodos disponíveis são componentes principais, mínimos quadrados não ponderados, mínimos quadrados generalizados, máxima verossimilhança, fatoração do eixo principal, fatoração alpha e fatoração de imagem.

- *Análise de Componentes Principais*. Um método de extração de fator usado para formar combinações lineares não correlacionadas das variáveis observadas. O primeiro componente possui variância máxima. Componentes sucessivos explicam progressivamente partes menores da variância e nenhum deles se correlaciona com o outro. A análise de componentes principais é utilizada para obter a solução de fator inicial. Ela pode ser usada quando uma matriz de correlações é singular.
- *Método de Quadrados Mínimos Não Ponderados*. Um método de extração de fator que minimiza a soma das diferenças quadráticas entre as matrizes de correlação observadas e reproduzidas (ignorando as diagonais).
- *Método de Quadrados Mínimos Generalizado*. Um método de extração de fator que minimiza a soma das diferenças quadradas entre as matrizes de correlação observadas e reproduzidas. As correlações são ponderadas pelo inverso de sua exclusividade, para que as variáveis com exclusividade alta recebam uma ponderação menor que aquelas com exclusividade baixa.
- *Método de Máxima Verossimilhança*. Um método de extração de fator que produz estimativas de parâmetros que provavelmente produzirão a matriz de correlações observadas se a amostra for de uma distribuição normal multivariada. As correlações são ponderadas pelo inverso da exclusividade das variáveis, e um algoritmo iterativo é utilizado.
- *Método de eixo principal de análise fatorial*. Um método de extrair fatores da matriz de correlação original, com coeficientes de correlação múltipla quadrados colocados na diagonal como estimativas iniciais das comunalidades. Esses carregamentos de fatores são utilizados para estimar novas comunalidades que substituem as estimativas de comunalidades antigas na diagonal. As iterações continuam até que as mudanças nas comunalidades de uma iteração para a próxima satisfaçam o critério de convergência para a extração.
- *Fatoração alfa*. Um método de extração de fatores que considera as variáveis na análise como uma amostra do universo de variáveis potenciais. Esse método maximiza a confiabilidade alfa dos fatores.
- *Método de imagem de análise fatorial*. Um método de extração de fatores desenvolvido por Guttman e baseado na teoria da imagem. A parte comum da variável, chamada de imagem parcial, é definida como sua regressão linear em variáveis restantes, ao invés de uma função de fatores hipotéticos.

**Analisar.** Permite especificar uma matriz de correlação ou uma matriz de correlações.

- **Matriz de correlações.** Útil se as variáveis em sua análise forem medidas em diferentes escalas.
- **Matriz de covariâncias.** Útil quando você deseja aplicar sua análise fatorial a vários grupos com diferentes variâncias para cada variável.

**Extrair.** É possível reter todos os fatores cujos autovalores excedem um valor especificado ou reter um número específico de fatores.

**Exibição.** Permite que você solicite a solução fatorial não rotacionada e um gráfico de escarpa dos autovalores.

- *Solução de Fator Não Girado*. Exibe carregamentos de fatores não girados (matriz padrão de fator), comunalidades e os autovalores para a solução de fator.
- *Gráfico de escarpa*. Um gráfico da variância que está associada a cada fator. Esse gráfico é utilizado para determinar quantos fatores devem ser mantidos. Normalmente, o gráfico mostra uma divisão distinta entre a inclinação íngreme dos fatores grandes e o acompanhamento gradual do restante (o scree).

**Máximo de Iterações para Convergência.** Permite especificar o número máximo de etapas que o algoritmo pode executar para estimar a solução.

## Rotação de Análise Fatorial

**Método.** Permite selecionar o método de rotação de fator e se aplicar a normalização do Kaiser. Os métodos disponíveis são varimax, oblimin direto, quartimax, equamax ou promax.

- *Método Varimax*. Um método de rotação ortogonal que minimiza o número de variáveis que possuem altos carregamentos em cada fator. Este método simplifica a interpretação dos fatores.

- *Método de Oblimin direto.* Um método para rotação oblíqua (não ortogonal). Quando delta é igual a 0 (o padrão), as soluções são mais oblíquas. Conforme delta se torna mais negativo, os fatores se tornam menos oblíquos. Para substituir o delta padrão de 0, insira um número menor ou igual a 0,8.
- *Método Quartimax.* Um método de rotação que minimiza o número de fatores necessários para explicar cada variável. Este método simplifica a interpretação das variáveis observadas.
- *Método Equamax.* Um método de rotação que é uma combinação do método varimax, que simplifica os fatores, e o método quartimax, que simplifica as variáveis. O número de variáveis que são altamente carregadas em um fator e o número de fatores necessários para explicar uma variável são minimizados.
- *Rotação Promax.* Uma rotação oblíqua, que permite que os fatores sejam correlacionados. Essa rotação pode ser calculada mais rapidamente do que uma rotação oblímim direta, sendo útil para conjuntos de dados grandes.
- *Aplicar a normalização Kaiser.* Por padrão, permite aplicar a normalização do Kaiser quando uma rotação é especificada.

**Exibição.** Permite incluir a saída na solução girada, bem como carregar gráficos para os primeiros dois ou três fatores.

- *Solução Girada.* Um método de rotação deve ser selecionado para obter uma solução girada. Para rotações ortogonais, a matriz de padrões girada e a matriz de transformação de fator são exibidas. Para rotações oblíquas, as matrizes de correlação de padrão, de estrutura e de fator são exibidas.
- *Gráfico de Carregamento de Fator.* Gráfico de carregamento de fatores tridimensionais dos três primeiros fatores. Para uma solução de dois fatores, um gráfico bidimensional é mostrado. O gráfico não será exibido se apenas um fator for extraído. Os gráficos exibem soluções giradas se a rotação for solicitada.

**Máximo de Iterações para Convergência.** Permite especificar o número máximo de passos que o algoritmo pode utilizar para executar a rotação.

## Escores de Análise Fatorial

**Salvar como variáveis.** Cria uma nova variável para cada fator na solução final.

**Método.** Os métodos alternativos para calcular os escores fatoriais são regressão, Bartlett e Anderson-Rubin. Anderson-Rubin.

- *Método de regressão.* Um método para estimar coeficientes de escore de fator. Os escores que são produzidos têm uma média de 0 e uma variância igual à correlação múltipla quadrada entre os escores dos fatores estimados e os valores de fatores reais. Os escores podem ser correlacionados mesmo quando fatores forem ortogonais.
- *Escores de Bartlett.* Um método de estimar coeficientes de escore de fator. As pontuações que são produzidas têm uma média de 0. A soma de quadrados dos fatores únicos sobre a gama de variáveis é minimizada.
- *Método de Anderson-Rubin.* Um método de estimar coeficientes de escore fatorial; uma modificação do método de Bartlett que assegura a ortogonalidade dos fatores estimados. Os escores que são produzidos possuem uma média de 0, um desvio padrão de 1 e são não correlacionados.

**Exibir matriz do coeficiente de escore dos fatores.** Mostra os coeficientes pelos quais as variáveis são multiplicadas para obter os escores dos fatores. Também mostra as correlações entre os escores dos fatores.

## Opções de Análise Fatorial

**Valores omissos.** Permite especificar como os valores ausentes são tratados. As opções disponíveis são excluir casos *listwise*, excluir casos *entre pares* ou substituir pela média.

**Formato de Exibição de Coeficiente.** Permite controlar aspectos das matrizes de saída. Você classifica os coeficientes por tamanho e suprime os coeficientes com valores absolutos menores que o valor especificado.



## Recursos Adicionais do Comando FACTOR

O idioma da sintaxe de comando também permite:

- Especifique os critérios de convergência para iteração durante a extração e rotação.
- Especifique gráficos de fator girado individuais.
- Especifique quantas pontuações de fator salvar.
- Especifique valores diagonais para o método de fatoraçoão do eixo principal.
- Grave matrizes de correlação ou matrizes de carregamento de fator em disco para análise posterior.
- Leia e analise matrizes de correlação ou matrizes de carregamento de fator.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Escolhendo um Procedimento para Clusterização

---

As análises de cluster podem ser realizadas usando o procedimento de análise TwoStep, Hierárquico ou K-médias. Cada procedimento emprega um algoritmo diferente para a criação de clusters, e cada um possui opções não disponíveis nos demais.

**Análise de Cluster TwoStep.** Para muitos aplicativos, o procedimento Análise de cluster TwoStep será o método de escolha. Ele fornece os recursos exclusivos a seguir:

- Seleção automática do melhor número de clusters, além de medidas para escolha entre os modelos de cluster.
- Capacidade de criar modelos de cluster simultaneamente com base em variáveis categóricas e contínuas.
- Capacidade de salvar o modelo de cluster em um arquivo XML externo e, em seguida, ler esse arquivo e atualizar o modelo de cluster usando dados mais recentes.

Além disso, o procedimento Análise de cluster TwoStep pode analisar arquivos de dados grandes.

**Análise de Cluster Hierárquica.** O procedimento Análise de cluster hierárquica é limitado a arquivos de dados menores (centenas de objetos a serem agrupados), mas possui os recursos exclusivos a seguir:

- Capacidade de agrupar casos ou variáveis.
- Capacidade de calcular uma variedade de soluções possíveis e salvar associações de cluster para cada uma dessas soluções.
- Vários métodos para formação de clusters, transformação de variáveis e medição da dissimilaridade entre clusters.

Contanto que todas as variáveis sejam do mesmo tipo, o procedimento Análise de Cluster Hierárquica pode analisar variáveis intervalares (contínuas), de contagem ou binárias.

**Análise de Cluster por K-Médias.** O procedimento Análise de cluster por k-médias é limitado a dados contínuos e exige que você especifique o número de clusters com antecedência, mas possui os recursos exclusivos a seguir:

- Capacidade de economizar distâncias de centros de cluster para cada objeto.
- Capacidade de ler centros do cluster iniciais e de salvar centros do cluster finais em um arquivo do IBM SPSS Statistics externo.

Além disso, o procedimento Análise de cluster por k-médias pode analisar grandes arquivos de dados.

## Análise de cluster TwoStep

---

O procedimento Análise de cluster TwoStep é uma ferramenta exploratória projetada para revelar agrupamentos naturais (ou clusters) dentro de um conjunto de dados que de outra forma não seriam aparentes. O algoritmo empregado por este procedimento possui várias características desejáveis que o diferenciam das técnicas tradicionais de agrupamento:

- **Manipulação de variáveis categóricas e contínuas.** Assumindo que as variáveis são independentes, uma distribuição multinomial-normal conjunta pode ser colocada em variáveis categóricas e contínuas.
- **Seleção automática do número de clusters.** Ao comparar os valores de um critério de escolha de modelo em diferentes soluções de agrupamento, o procedimento pode determinar automaticamente o número ideal de agrupamentos.
- **Escalabilidade.** Ao construir uma árvore de recursos de cluster (CF) que resume os registros, o algoritmo TwoStep permite analisar arquivos de dados grandes.

**Exemplo.** As empresas de varejo e de produtos do consumidor aplicam regularmente técnicas de armazenamento em cluster a dados que descrevem os hábitos de compra, sexo, idade, nível de renda, etc. de seus clientes. Essas empresas customizam suas estratégias de desenvolvimento de marketing e de produtos para aumentar as vendas e criar fidelidade à marca.

**Medida de distância.** Essa seleção determina como a similaridade entre dois clusters é calculada.

- **Log da verossimilhança.** A medida de probabilidade coloca uma distribuição de probabilidade nas variáveis. Variáveis contínuas são consideradas como sendo distribuídas normalmente, ao passo que as variáveis categóricas são consideradas como sendo multinomiais. Todas as variáveis são consideradas independentes.
- **Euclidiana.** A medida Euclidiana é a distância em "linha reta" entre dois clusters. Ele pode ser usado somente quando todas as variáveis são contínuas.

**Número de Clusters.** Essa seleção permite especificar como o número de clusters deve ser determinado.

- **Determinar automaticamente.** O procedimento determinará automaticamente o "melhor" número de clusters, usando o critério especificado no grupo Clustering Criterion. Opcionalmente, insira um número inteiro positivo especificando o número máximo de clusters que o procedimento deve considerar.
- **Especificar fixo.** Permite corrigir o número de clusters na solução. Insira um número inteiro positivo.

**Contagem de Variáveis Contínuas.** Este grupo fornece um resumo das especificações de padronização de variáveis contínuas feitas na caixa de diálogo Opções. Consulte o tópico [“Opções de Análise de Cluster TwoStep”](#) na página 187 para obter mais informações

**Critério de clusterização.** Esta seleção determina como o algoritmo de armazenamento em cluster automático determina o número de clusters. O Critério de Informação Bayesiano (BIC) ou o Akaike Information Criterion (AIC) podem ser especificados.

Considerações sobre dados de análise de cluster TwoStep

**Dados.** Este procedimento funciona com variáveis contínuas e categóricas. Casos representam objetos a serem agrupados e as variáveis representam atributos nos quais o agrupamento é baseado.

**Ordem de Caso.** Observe que a árvore de recursos do cluster e a solução final podem depender da ordem dos casos. Para minimizar os efeitos da ordem, ordene os casos de forma aleatória. Você pode querer obter várias soluções diferentes com casos classificados em diferentes ordens aleatórios para verificar a estabilidade de uma determinada solução. Em situações em que isso é difícil devido a tamanhos de arquivo extremamente grandes, várias execuções com uma amostra de casos classificados em diferentes ordens aleatórios podem ser substituídas.

**Suposições.** A medida de distância de verossimilhança assume que as variáveis no modelo de cluster são independentes. Além disso, supõe-se que cada variável contínua tenha uma distribuição normal (Gaussiana) e que cada variável categórica tenha uma distribuição multinomial. Testes internos empíricos indicam que o procedimento é bastante robusto a violações tanto da suposição de independência quanto das suposições distributivas, mas você deve tentar estar ciente de quão bem essas suposições são atendidas.

Utilize o procedimento [Correlações Bivariadas](#) para testar a independência das duas variáveis contínuas. Utilize o procedimento [Crosstabs](#) para testar a independência das duas variáveis categóricas. Utilize o procedimento [Médias](#) para testar a independência entre uma variável contínua e uma variável categórica. Utilize o procedimento [Explorar](#) para testar a normalidade de uma variável contínua. Utilize o procedimento [Teste Qui-Quadrado](#) para testar se uma variável categórica possui uma distribuição multinomial especificada.

Para obter uma análise de cluster TwoStep

1. A partir dos menus, escolha:

**Analisar > Classificar > Cluster TwoStep...**

2. Selecione uma ou mais variáveis categóricas ou contínuas.

Opcionalmente, é possível:

- Ajuste os critérios pelos quais os clusters são construídos.
- Selecione as configurações para manipulação de ruído, alocação de memória, padronização de variáveis e entrada de modelo de cluster.
- Solicitar a saída do visualizador de modelo.
- Salve os resultados do modelo no arquivo de trabalho ou em um arquivo XML externo.

## Opções de Análise de Cluster TwoStep

**Tratamento de Valor Discrepante.** Este grupo permite tratar valores discrepantes especialmente durante clusterização se a árvore de recursos de cluster (CF) for preenchida. A árvore CF estará cheia se ela não puder aceitar mais nenhum caso em um nó folha e nenhum nó folha puder ser dividido.

- Se selecionar tratamento de ruído e a árvore CF for preenchida, a árvore crescerá novamente após colocar os casos nas folhas esparsas em uma folha de "ruído". Uma folha será considerada esparsa se ela contiver menos do que a porcentagem especificada de casos do tamanho máximo de folha. Após a árvore ter crescido novamente, os valores discrepantes serão colocados na árvore CF, se possível. Caso contrário, os valores discrepantes serão descartados.
- Se você não selecionar tratamento de ruído e a árvore CF for preenchida, ela crescerá novamente utilizando um limite de mudança de distância maior. Após a clusterização final, os valores que não puderem ser designados a um cluster serão rotulados como valores discrepantes. O cluster de valor discrepante recebe um número de identificação de -1 e não é incluído na contagem do número de clusters.

**Alocação de Memória.** Este grupo permite especificar a quantidade máxima de memória em megabytes (MB) que o algoritmo de cluster deve usar. Se o procedimento ultrapassar esse máximo, ele usará o disco para armazenar informações que não caberão na memória. Especifique um número maior ou igual a 4.

- Consulte o administrador do sistema para obter o maior valor que pode ser especificado em seu sistema.
- O algoritmo poderá falhar ao localizar o número correto ou especificado de clusters se esse valor for muito baixo.

**Padronização de variável.** O algoritmo de clusterização funciona com variáveis contínuas padronizadas. Quaisquer variáveis contínuas que não estiverem padronizadas devem ser deixadas como variáveis na lista A ser padronizada. Para economizar tempo e esforço computacional, é possível selecionar quaisquer variáveis contínuas que já tiverem sido padronizadas como variáveis na lista Padronizadas Assumidas.

Opções avançadas

**Critérios de Ajuste de Árvore CF.** As configurações do algoritmo de clustering a seguir se aplicam especificamente à árvore de recursos de cluster (CF) e devem ser alteradas com cuidado:

- **Limite de Mudança da Distância Inicial.** Este é o limite inicial usado para crescer a árvore CF. Se inserir um determinado caso em uma folha da árvore CF gerar uma tensão menor que o limite, a folha não será dividida. Se a tensão exceder o limite, a folha é dividida.
- **Máximo de ramificações (por nó folha).** O número máximo de nós filhos que um nó folha pode ter.
- **Profundidade Máxima da Árvore.** O número máximo de níveis que a árvore CF pode ter.
- **Número Máximo de Nós Possíveis.** Isso indica o número máximo de nós de árvore CF que podem potencialmente ser gerados pelo procedimento, com base na função  $(b^{d+1} - 1) / (b - 1)$ , em que  $b$  é o máximo de ramificações e  $d$  é a profundidade máxima da árvore. Lembre-se de que sobrepor uma

árvore CF grande poderá esgotar os recursos do sistema e prejudicar o desempenho do procedimento. No mínimo, cada nó requer 16 bytes.

**Atualização do Modelo de Cluster.** Este grupo permite importar e atualizar um modelo de cluster gerado em uma análise anterior. O arquivo de entrada contém a árvore CF em formato XML. Em seguida, o modelo será atualizado com os dados no arquivo ativo. Os nomes de variáveis devem ser selecionados na caixa de diálogo principal na mesma ordem em que eles foram especificados na análise anterior. O arquivo XML permanece inalterado, a menos que você grave especificamente as informações do novo modelo no mesmo nome de arquivo. Consulte o tópico [“Saída da Análise de Cluster TwoStep” na página 188](#) para obter mais informações

Se uma atualização do modelo de cluster for especificada, as opções referentes à geração da árvore CF que foram especificadas para o modelo original serão utilizadas. Mais especificamente, as configurações de critérios de ajuste de medida de distância, de tratamento de ruído, de alocação de memória ou de árvore CF para o modelo salvo são utilizadas, e todas as configurações para essas opções nas caixas de diálogo são ignoradas.

*Nota:* Ao executar uma atualização do modelo de cluster, o procedimento supõe que nenhum dos casos selecionados no conjunto de dados ativo foi utilizado para criar o modelo de cluster original. O procedimento também supõe que os casos usados na atualização do modelo são provenientes da mesma população que os casos utilizados para criar o modelo original, ou seja, as médias e variâncias de variáveis contínuas e os níveis de variáveis categóricas são considerados os mesmos nos dois conjuntos de casos. Se os seus conjuntos de casos "novos" e "antigos" forem provenientes de populações heterogêneas, deve-se executar o procedimento de Análise de Cluster TwoStep nos conjuntos de casos combinados para obter os melhores resultados.

## Saída da Análise de Cluster TwoStep

**Saída.** Este grupo fornece opções para exibir os resultados de clusterização.

- **Tabelas dinâmicas.** Os resultados são exibidos em tabelas dinâmicas.
- **Gráficos e tabelas no Visualizador de Modelo.** Os resultados são exibidos no Visualizador de Modelo.
- **Campos de avaliação.** Isso calcula os dados do cluster para as variáveis que não forem utilizadas na criação do cluster. Campos de avaliação podem ser exibidos ao longo dos recursos de entrada no visualizador de modelo ao selecioná-los no subdiálogo Exibir. Campos com valores omissos são ignorados.

**Trabalhando com Arquivo de Dados.** Este grupo permite salvar as variáveis no conjunto de dados ativo.

- **Criar variável de associação de cluster.** Esta variável contém um número de identificação do cluster para cada caso. O nome dessa variável é *tsc\_n*, em que *n* é um número inteiro positivo indicando o ordinal da operação de salvamento do conjunto de dados ativo concluída por este procedimento em uma determinada sessão.

**Arquivos XML.** O modelo de cluster final e a árvore CF são dois tipos de arquivos de saída que podem ser exportados em formato XML.

- **Exportar modelo final.** O modelo de cluster final é exportado no arquivo especificado em formato XML (PMML). É possível usar esse arquivo de modelo para aplicar as informações de modelo a outros arquivos de dados para propósitos de escoragem. para obter mais informações.
- **Exportar árvore CF.** Esta opção permite salvar o estado atual da árvore de cluster e atualizá-la posteriormente usando dados mais recentes.

## O visualizador de cluster

Os modelos de cluster são normalmente usados para encontrar grupos (ou clusters) de registros semelhantes com base nas variáveis examinadas, onde a similaridade entre membros de um mesmo grupo é alta e a similaridade entre membros de grupos diferentes é baixa. Os resultados podem ser usados para identificar associações que de outra forma não seriam aparentes. Por exemplo, por meio da análise de cluster das preferências dos clientes, nível de renda e hábitos de compra, pode ser possível

identificar os tipos de clientes com maior probabilidade de responder a uma determinada campanha de marketing.

Há duas abordagens para interpretar os resultados em uma exibição de cluster:

- Examine os clusters para determinar as características exclusivas desse cluster. *Um cluster contém todos os solicitantes de crédito de alta renda? Esse cluster contém mais registros do que os outros?*
- Examine os campos nos clusters para determinar como os valores são distribuídos entre os clusters. *O nível de educação de uma pessoa determina a associação em um cluster? Um escore de risco de crédito alto se distingue entre associação em um cluster ou outro?*

Usando as visualizações principais e as várias visualizações vinculadas no Visualizador de cluster, é possível obter insights para ajudá-lo a responder a essas perguntas.

Para ver informações sobre o modelo de cluster, ative (dê um clique duplo) o objeto Visualizador de modelo no Visualizador .

## Visualizador de Cluster

O Visualizador de Cluster é formado por dois painéis, a visualização principal à esquerda e a visualização vinculada, ou auxiliar, à direita. Existem duas visualizações principais:

- Sumarização do Modelo (o padrão). Consulte o tópico [“Visualização de Sumarização do Modelo” na página 189](#) para obter mais informações
- Clusters. Consulte o tópico [“Visualização Clusters” na página 190](#) para obter mais informações

Há quatro visualizações vinculadas/auxiliares:

- Importância do Preditor. Consulte o tópico [“Visualização de Importância do Preditor de Cluster” na página 191](#) para obter mais informações
- Tamanhos do Cluster (o padrão). Consulte o tópico [“Visualização de tamanhos de cluster” na página 191](#) para obter mais informações
- Distribuição de célula. Consulte o tópico [“visualização de distribuição de células” na página 192](#) para obter mais informações
- Comparação de cluster. Consulte o tópico [“Visualização de comparação de cluster” na página 192](#) para obter mais informações

## Visualização de Sumarização do Modelo

A visualização Resumo do Modelo mostra um instantâneo, ou resumo, do modelo de cluster, incluindo uma medida Silhouette de coesão e separação de cluster que é sombreada para indicar resultados ruins, justos ou bons. Essa captura instantânea permite verificar rapidamente se a qualidade é ruim; nesse caso, você pode decidir retornar ao nó de modelagem para alterar as configurações do modelo de cluster para produzir um resultado melhor.

The results of poor, fair, and good are based on the work of Kaufman and Rousseeuw (1990) regarding interpretation of cluster structures. Na visão do Resumo do Modelo, um bom resultado equivale a dados que refletem a classificação de Kaufman e Rousseeuw como evidência razoável ou forte da estrutura do cluster, razoável reflete sua classificação de evidência fraca e ruim reflete sua classificação de nenhuma evidência significativa.

A silhueta mede médias, sobre todos os registros,  $(B-A) / \max(A,B)$ , onde A é a distância do registro ao centro do cluster e B é a distância do registro ao centro do cluster mais próximo ao qual não pertence. Um coeficiente de silhueta de 1 significaria que todos os casos estão localizados diretamente em seus centros de cluster. Um valor de -1 significaria que todos os casos estão localizados nos centros de cluster de algum outro cluster. Um valor de 0 significa, em média, que os casos são equidistantes entre seu próprio centro de cluster e o outro cluster mais próximo.

O resumo inclui uma tabela que contém as informações a seguir:

- **Algoritmo.** O algoritmo de agrupamento usado, por exemplo, "TwoStep".
- **Variáveis de Entrada.** O número de campos, também conhecido como **entradas** ou **preditores**.

- **Clusters.** O número de clusters na solução.

## **Visualização Clusters**

A visualização Clusters contém uma grade cluster por recursos que inclui nomes, tamanhos e perfis de cluster para cada cluster.

As colunas na grade contêm as informações a seguir:

- **Cluster.** Os números de cluster criados pelo algoritmo.
- **Rótulo.** Quaisquer rótulos aplicados a cada cluster (em branco por padrão). Dê um clique duplo na célula para inserir um rótulo que descreva o conteúdo do cluster; por exemplo, "compradores de carros de luxo".
- **Descrição.** Qualquer descrição do conteúdo do cluster (em branco por padrão). Clique duas vezes na célula para inserir uma descrição do cluster; por exemplo, "mais de 55 anos, profissionais, ganhando mais de US\$ 100.000".
- **Tamanho.** O tamanho de cada cluster como uma porcentagem da amostra geral de clusters. Cada célula de tamanho na grade exibe uma barra vertical que mostra a porcentagem de tamanho dentro do cluster, uma porcentagem de tamanho em formato numérico e as contagens de maiúsculas e minúsculas do cluster.
- **Variáveis.** As entradas individuais ou preditores, classificados por importância geral por padrão. Se alguma coluna tiver tamanhos iguais, ela será mostrada em ordem crescente dos números do cluster.

A importância geral do recurso é indicada pela cor do sombreado do plano de fundo da célula; a característica mais importante é a mais escura; a característica menos importante é não sombreada. Um guia acima da tabela indica a importância atribuída a cada cor de célula de recurso.

Quando você passa o mouse sobre uma célula, o nome/rótulo completo do recurso e o valor de importância para a célula são exibidos. Mais informações podem ser exibidas, dependendo da visualização e do tipo de recurso. Na visualização Centros de clusters, isso inclui a estatística da célula e o valor da célula; por exemplo: "Média: 4,32". Para características categóricas, a célula mostra o nome da categoria (modal) mais frequente e sua porcentagem.

Na visualização Clusters, é possível selecionar várias maneiras de exibir as informações do cluster:

- Transponha clusters e recursos. Consulte o tópico [“Transpor Clusters e Variáveis”](#) na página 190 para obter mais informações.
- Classifique recursos. Consulte o tópico [“Ordenar Variáveis”](#) na página 190 para obter mais informações.
- Ordenar clusters. Consulte o tópico [“Ordenar Clusters”](#) na página 191 para obter mais informações.
- Selecione o conteúdo da célula. Consulte o tópico [“Conteúdo das células”](#) na página 191 para obter mais informações.

### *Transpor Clusters e Variáveis*

Por padrão, os clusters são exibidos como colunas e os recursos são exibidos como linhas. Para reverter essa exibição, clique no botão **Transpor Clusters e Variáveis** à esquerda dos botões **Classificar Variáveis Por**. Por exemplo, você pode querer fazer isso quando tiver muitos clusters exibidos, para reduzir a quantidade de rolagem horizontal necessária para ver os dados.

### *Ordenar Variáveis*

Os botões **Ordenar Variáveis Por** permitem selecionar como as células de variáveis são exibidas:

- **Importância Geral.** Esse é o padrão de ordenação. Os recursos são classificados em ordem decrescente de importância geral e a ordem de classificação é a mesma nos clusters. Se algum recurso tiver valores de importância vinculados, os recursos vinculados serão listados em ordem de classificação crescente dos nomes dos recursos.
- **Importância Dentro do Cluster.** Os recursos são classificados de acordo com sua importância para cada cluster. Se algum recurso tiver valores de importância vinculados, os recursos vinculados serão

listados em ordem de classificação crescente dos nomes dos recursos. Quando essa opção é escolhida, a ordem de classificação geralmente varia entre os clusters.

- **Nome.** Os recursos são classificados por nome em ordem alfabética.
- **Ordem de dados.** Os recursos são classificados por ordem no conjunto de dados.

#### *Ordenar Clusters*

Por padrão, os clusters são classificados em ordem decrescente de tamanho. Os botões **Ordenar Clusters Por** permitem classificá-los por nome em ordem alfabética ou, se você tiver criado rótulos exclusivos, em ordem alfanumérica de rótulo.

Os recursos que têm o mesmo rótulo são classificados pelo nome do cluster. Se os clusters forem classificados por rótulo e você editar o rótulo de um cluster, a ordem de classificação será atualizada automaticamente.

#### *Conteúdo das células*

Os botões de **Células** lhe permitem mudar a exibição dos conteúdos da célula para variáveis e campos de avaliação.

- **Centros do Cluster.** Por padrão, as células exibem nomes/rótulos de recursos e a tendência central para cada combinação de cluster/recurso. A média é mostrada para campos contínuos e o modo (categoria de ocorrência mais frequente) com porcentagem de categoria para campos categóricos.
- **Distribuições Absolutas.** Mostra nomes/rótulos de recursos e distribuições absolutas dos recursos em cada cluster. Para recursos categóricos, a exibição mostra gráficos de barras sobrepostos com categorias ordenadas em ordem crescente dos valores de dados. Para recursos contínuos, a exibição mostra um gráfico de densidade suave que usa os mesmos pontos finais e intervalos para cada cluster.

A exibição de cor vermelha sólida mostra a distribuição do cluster, enquanto a exibição mais pálida representa os dados gerais.

- **Distribuições Relativas.** Mostra nomes/rótulos de recursos e distribuições relativas nas células. Em geral, as exibições são semelhantes às exibidas para distribuições absolutas, exceto que as distribuições relativas são exibidas.

A exibição de cor vermelha sólida mostra a distribuição do cluster, enquanto a exibição mais clara representa os dados gerais.

- **Visualização Básica.** Onde há muitos clusters, pode ser difícil ver todos os detalhes sem rolar. Para reduzir a quantidade de rolagem, selecione esta exibição para alterar a exibição para uma versão mais compacta da tabela.

### ***Visualização de Importância do Preditor de Cluster***

A visualização de importância do preditor mostra a importância relativa de cada campo na estimativa do modelo.

### ***Visualização de tamanhos de cluster***

A visualização Tamanhos de Cluster mostra um gráfico de pizza que contém cada cluster. O tamanho percentual de cada cluster é mostrado em cada fatia; passe o mouse sobre cada fatia para exibir a contagem nessa fatia.

Abaixo do gráfico, uma tabela lista as informações de tamanho a seguir:

- O tamanho do menor cluster (uma contagem e uma porcentagem do todo).
- O tamanho do maior cluster (tanto uma contagem quanto uma porcentagem do todo).
- A razão do tamanho do maior cluster para o menor cluster.

## visualização de distribuição de células

A visualização Distribuição de Células mostra um plot expandido e mais detalhado da distribuição dos dados de qualquer célula de variável que você seleciona na tabela no painel principal Clusters.

## Visualização de comparação de cluster

A visualização de Comparação de Cluster consiste em um layout de estilo em grade, com variáveis nas linhas e clusters selecionados nas colunas. Essa visualização ajuda a entender melhor os fatores que compõem os clusters e também permite ver diferenças entre os clusters não apenas comparado com dados gerais, mas também entre si.

Para selecionar clusters para exibição, clique na parte superior da coluna de cluster no painel principal de Clusters. Use Ctrl-clique ou Shift-clique para selecionar ou cancelar a seleção de mais de um cluster para comparação.

*Nota:* É possível selecionar até cinco clusters para exibição.

Os clusters são mostrados na ordem em que eles foram selecionados, ao passo que a ordem dos campos é determinada pela opção **Classificar Variáveis Por**. Ao selecionar **Importância Dentro do Cluster**, os campos são sempre ordenados por importância geral.

Os gráficos em segundo plano mostram as distribuições gerais de cada variável:

- As variáveis categóricas são mostradas como gráficos de ponto, em que o tamanho do ponto indica a categoria mais frequente/modal para cada cluster (por variável).
- Variáveis contínuas são exibidas como boxplots, que mostram medianas gerais e as amplitudes interquartis.

Sobrepostos nestas visualizações em segundo plano estão os boxplots para clusters selecionados:

- Para variáveis contínuas, os marcadores de ponto quadrados e as linhas horizontais indicam a mediana e a amplitude interquartil para cada cluster.
- Cada cluster é representado por uma cor diferente, mostrada na parte superior da visualização.

## Navegando no Visualizador de Cluster

O Visualizador de Cluster é uma exibição interativa. Será possível:

- Selecione um campo ou cluster para visualizar mais detalhes.
- Compare clusters para selecionar itens de interesse.
- Altere a exibição.
- Transpor eixos.

Usando as barras de ferramentas

Você controla as informações mostradas nos painéis esquerdo e direito usando as opções da barra de ferramentas. É possível mudar a orientação da exibição (de cima para baixo, da esquerda para a direita ou da direita para a esquerda) usando os controles da barra de ferramentas. Além disso, também é possível redefinir o visualizador para as configurações padrão e abrir uma caixa de diálogo para especificar o conteúdo da visualização Clusters no painel principal.

As opções **Ordenar Variáveis Por**, **Ordenar Clusters Por**, **Células** e **Exibir** estão disponíveis somente quando você seleciona a visualização de **Clusters** no painel principal. Consulte o tópico “[Visualização Clusters](#)” na página 190 para obter mais informações.


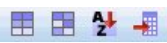
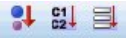

Ícone	Tópico
	Consulte <a href="#">Transpor Clusters e Variáveis</a>



Tabela 2. Ícones da barra de ferramentas (continuação)

Ícone	Tópico
	Consulte <a href="#">Ordenar Variáveis Por</a>
	Consulte <a href="#">Ordenar Clusters Por</a>
	Consulte <a href="#">Células</a>

Exibição de visualização de cluster de controle

Para controlar o que é mostrado na visualização de Clusters no painel principal, clique no botão **Exibir**; o diálogo Exibir é aberto.

**Variáveis.** Selecionado por padrão. Para ocultar todas as variáveis de entrada, desmarque a caixa de seleção.

**Campos de Avaliação.** Escolha os campos de avaliação (campos não utilizados para criar o modelo de cluster, mas enviados para o visualizador de modelos para avaliar os clusters) para exibição; nenhum é mostrado por padrão. *Nota* o campo de avaliação deve ser uma sequência com mais de um valor. Esta caixa de seleção fica indisponível se nenhum campo de avaliação estiver disponível.

**Descrições de Cluster.** Selecionado por padrão. Para ocultar todas as células de descrição do cluster, cancele a seleção da caixa de seleção.

**Tamanhos de Cluster.** Selecionado por padrão. Para ocultar todas as células de tamanho de cluster, desmarque a caixa de seleção.

**Número Máximo de Categorias.** Especifique o número máximo de categorias a serem exibidas em gráficos de recursos categóricos; o padrão é 20.

## Filtrando Registros

Se você quiser saber mais sobre os casos em um determinado cluster ou grupo de clusters, poderá selecionar um subconjunto de registros para análise posterior com base nos clusters selecionados.

1. Selecione os clusters na visualização de Cluster do Visualizador de Cluster. Para selecionar vários clusters, use Ctrl-clique
2. A partir dos menus, escolha:
  - Gerar > Filtrar Registros...**
3. Insira um nome de variável de filtro. Os registros dos clusters selecionados receberão o valor 1 para este campo. Todos os outros registros receberão o valor 0 e serão excluídos das análises subsequentes até que você mude o status do filtro.
4. Clique em **OK**.

## Análise de cluster hierárquica

Este procedimento tenta identificar grupos relativamente homogêneos de casos (ou variáveis) com base em características selecionadas, usando um algoritmo que começa com cada caso (ou variável) em um cluster separado e combina clusters até restar apenas um. É possível analisar variáveis brutas, ou escolher a partir de uma variedade de transformações de padronização. Medidas de distância ou similaridade são geradas pelo procedimento de proximidades. As estatísticas são exibidas em cada estágio para ajudá-lo a selecionar a melhor solução.

**Exemplo.** Existem grupos identificáveis de programas de televisão que atraem audiências semelhantes dentro de cada grupo? Com a análise de cluster hierárquico, é possível agrupar programas de televisão (casos) em grupos homogêneos com base nas características do espectador. Isso pode ser usado para identificar segmentos para marketing. Ou é possível agrupar cidades (casos) em grupos homogêneos para que cidades comparáveis possam ser selecionadas para testar várias estratégias de marketing.

**Estatísticas.** Planejamento de aglomeração, matriz de distância (ou similaridade) e associação de cluster para uma única solução ou um intervalo de soluções. Gráficos: dendrogramas e gráficos icicle.

Considerações de Dados de Análise de Cluster Hierárquica

**Dados.** As variáveis podem ser dados quantitativos, binários ou de contagem. O dimensionamento de variáveis é uma questão importante - as diferenças no dimensionamento podem afetar sua(s) solução(ões) de cluster. Se suas variáveis tiverem grandes diferenças de escala (por exemplo, uma variável é medida em dólares e a outra é medida em anos), você deve considerar padronizá-las (isso pode ser feito automaticamente pelo procedimento Hierarchical Cluster Analysis).

**Ordem de caso.** Se existirem distâncias ou semelhanças vinculadas nos dados de entrada ou ocorrerem entre clusters atualizados durante a junção, a solução de cluster resultante pode depender da ordem dos casos no arquivo. Você pode querer obter várias soluções diferentes com casos classificados em diferentes ordens aleatórias para verificar a estabilidade de uma determinada solução.

**Suposições.** As medidas de distância ou similaridade utilizadas devem ser apropriadas para os dados analisados (veja o procedimento Proximidades para mais informações sobre opções de medidas de distância e similaridade). Além disso, deve-se incluir todas as variáveis relevantes em sua análise. A omissão de variáveis influentes pode resultar em uma solução enganosa. Como a análise de agrupamento hierárquico é um método exploratório, os resultados devem ser tratados como provisórios até que sejam confirmados com uma amostra independente.

Para obter uma análise de cluster hierárquico

1. No menu, escolha:

**Analisar > Classificar > Cluster Hierárquico...**

2. Se você estiver agrupando casos, selecione pelo menos uma variável numérica. Se estiver armazenando variáveis em cluster, selecione pelo menos três variáveis numéricas.

Opcionalmente, é possível selecionar uma variável de identificação para rotular casos.

## Método de Análise de Cluster Hierárquica

**Método de Cluster.** As alternativas disponíveis são ligação entre grupos, ligação dentro de grupos, vizinho mais próximo, vizinho mais distante, agrupamento centroide, agrupamento mediano e o método de Ward.

**Medida.** Permite especificar a medida de distância ou similaridade a ser usada no agrupamento.

Selecione o tipo de dados e a medida de distância ou similaridade apropriada:

- **Intervalo.** As alternativas disponíveis são distância euclidiana, distância euclidiana quadrada, cosseno, correlação de Pearson, Chebychev, bloco, Minkowski e customizado.
- **Contagens.** As alternativas disponíveis são a medida qui-quadrado e a medida fi-quadrado.
- **Binário.** As alternativas disponíveis são distância euclidiana, distância euclidiana quadrática, diferença de tamanho, diferença padrão, variância, dispersão, forma, correspondência simples, correlação fi de 4 pontos, lambda,  $D$  de Anderberg, divisão, Hamann, Jaccard, Kulczynski 1, Kulczynski 2, Lance e Williams, Ochiai, Rogers e Tanimoto, Russel e Rao, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Sokal e Sneath 4, Sokal e Sneath 5,  $Y$  de Yule e  $Q$  de Yule.

**Transformar Valores.** Permite padronizar valores de dados para casos ou valores antes de calcular as proximidades (não disponível para dados binários). Os métodos de padronização disponíveis são escores  $z$ , intervalo -1 a 1, intervalo 0 a 1, magnitude máxima de 1, média de 1 e desvio padrão de 1.

**Transformar Medidas.** Permite transformar os valores gerados pela medida de distância. Eles são aplicados após a medida de distância ter sido calculada. As alternativas disponíveis são valores absolutos, alterar o sinal e redimensionar para o intervalo de 0–1.

## Estatísticas de Análise de Cluster Hierárquica

**Planejamento de aglomeração.** Exibe os casos ou clusters combinados em cada estágio, as distâncias entre os casos ou clusters que estão sendo combinados e o último nível do cluster no qual um caso (ou variável) se uniu ao cluster.

**Matriz de proximidade.** Fornece as distâncias ou semelhanças entre os itens.

**Associação de Cluster.** Exibe o cluster ao qual cada caso é atribuído em um ou mais estágios na combinação de clusters. As opções disponíveis são solução única e gama de soluções.

## Gráficos de Análise de Cluster Hierárquica

**Dendrograma.** Exibe um *dendrograma*. Dendrogramas podem ser utilizados para avaliar a coesão dos clusters formados e pode fornecer informações sobre o número apropriado de clusters a serem mantidos.

**Icicle.** Exibe um *gráfico icicle*, incluindo todos os clusters ou um intervalo de clusters especificado. Os gráficos icicle exibem informações sobre como os casos são combinados em clusters em cada iteração da análise. A orientação permite selecionar um gráfico vertical ou horizontal.

## Análise de cluster hierárquica Salvar novas variáveis

**Associação de Cluster.** Permite que você salve associações de cluster para uma única solução ou uma variedade de soluções. As variáveis salvas podem, então, ser utilizadas em análises subsequentes para explorar outras diferenças entre os grupos.

## Recursos Adicionais da Sintaxe do Comando CLUSTER

O procedimento Cluster Hierárquico utiliza a sintaxe de comando do CLUSTER. O idioma da sintaxe de comando também permite:

- Utilizar vários métodos de clusterização em uma análise única.
- Ler e analisar uma matriz de proximidade.
- Gravar uma matriz de proximidade no disco para análise posterior.
- Especificar quaisquer valores para potência e raiz na medida de distância customizada (Potência).
- Especificar nomes para variáveis salvas.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Análise de Cluster por K-Médias

---

Este procedimento tenta identificar grupos de casos relativamente homogêneos com base em características selecionadas, usando um algoritmo que pode lidar com um grande número de casos. No entanto, o algoritmo exige que você especifique o número de clusters. É possível especificar centros de cluster iniciais se souber dessas informações. É possível selecionar um dos dois métodos para classificar casos, atualizando os centros de cluster de forma iterativa ou classificando apenas. É possível salvar a associação do cluster, as informações de distância e os centros finais do cluster. Opcionalmente, é possível especificar uma variável cujos valores são usados para rotular a saída caso a caso. Também é possível solicitar estatísticas  $F$  de análise de variância. Embora essas estatísticas sejam oportunistas (o procedimento tenta formar grupos que diferem), o tamanho relativo das estatísticas fornece informações sobre a contribuição de cada variável para a separação dos grupos.

**Exemplo.** Quais são alguns grupos identificáveis de programas de televisão que atraem audiências semelhantes dentro de cada grupo? Com a análise de cluster por  $k$ -médias, é possível agrupar programas de televisão (casos) em  $k$  grupos homogêneos com base nas características do telespectador. Esse processo pode ser usado para identificar segmentos para marketing. Ou é possível agrupar cidades (casos) em grupos homogêneos para que cidades comparáveis possam ser selecionadas para testar várias estratégias de marketing.

**Estatísticas.** Solução completa: centros do cluster iniciais, tabela ANOVA. Cada caso: informações do cluster, distância do centro do cluster.

Considerações de Dados de Análise de Cluster por K-Médias

**Dados.** As variáveis devem ser quantitativas no nível de intervalo ou razão. Se suas variáveis forem binárias ou contagens, use o procedimento de Análise de Cluster Hierárquica.

**Caso e ordem inicial do centro do cluster.** O algoritmo padrão para escolher os centros de cluster iniciais não é invariável à ordenação de casos. A opção **Utilizar médias de execução** na caixa de diálogo Iterar torna a solução resultante potencialmente dependente da ordem do caso, não importa como os centros do cluster iniciais forem escolhidos. Se você estiver usando um desses métodos, talvez queira obter várias soluções diferentes com casos classificados em diferentes ordens aleatórias para verificar a estabilidade de uma determinada solução. Especificar centros do cluster iniciais e não utilizar a opção **Utilizar médias de execução** evitam problemas relacionados à ordem do caso. No entanto, a ordenação dos centros de cluster iniciais pode afetar a solução se houver distâncias vinculadas dos casos aos centros de cluster. Para avaliar a estabilidade de uma determinada solução, é possível comparar resultados de análises com diferentes permutações dos valores iniciais do centro.

**Suposições.** As distâncias são calculadas usando a distância euclidiana simples. Se você quiser usar outra medida de distância ou similaridade, use o procedimento de Análise de Cluster Hierárquica. O ajuste de escala de variáveis é uma consideração importante. Se suas variáveis são medidas em escalas diferentes (por exemplo, uma variável é expressa em dólares e outra variável é expressa em anos), seus resultados podem ser enganosos. Nesses casos, deve-se considerar a padronização de suas variáveis antes de executar a análise de cluster por  $k$ -médias (esta tarefa pode ser feita no procedimento Descritivos). O procedimento pressupõe que você selecionou o número apropriado de clusters e que incluiu todas as variáveis relevantes. Se você escolheu um número inadequado de clusters ou omitiu variáveis importantes, seus resultados podem ser enganosos.

Para obter uma análise de cluster K-médias

1. A partir dos menus, escolha:

**Analisar > Classificar > Cluster de K-Médias...**

2. Selecione as variáveis a serem usadas na análise de cluster.

3. Especifique o número de clusters. (O número de clusters deve ser pelo menos 2 e não deve ser maior que o número de casos no arquivo de dados.)

4. Selecione **Iterar e classificar** ou **Somente classificar**.

5. Opcionalmente, selecione uma variável de identificação para rotular os casos.

## Eficiência de Análise de Cluster por K-Médias

O comando de análise de cluster por  $k$ -médias é eficaz principalmente porque ele não calcula as distâncias entre todos os pares de casos, assim como fazem muitos algoritmos de cluster, incluindo o algoritmo que é utilizado pelo comando de clusterização hierárquico.

Para obter o máximo de eficiência, pegue uma amostra de casos e selecione o método **Iterar e classificar** para determinar os centros do cluster. Selecione **Gravar final como**. Em seguida, restaure o arquivo de dados inteiro e selecione **Somente classificar** como o método e selecione **Leitura inicial de** para classificar o arquivo inteiro usando os centros que são estimados a partir da amostra. É possível gravar e ler a partir de um arquivo ou de um conjunto de dados. Conjuntos de dados estão disponíveis para uso subsequente na mesma sessão, mas não são salvos como arquivos, a menos que sejam salvos explicitamente antes do término da sessão. Os nomes do conjunto de dados devem estar de acordo com as regras de nomenclatura de variáveis. Veja o tópico para obter mais informações.

## Iteração de Análise de Cluster por K-Médias

*Nota:* Estas opções estarão disponíveis apenas se selecionar o método **Iterar e classificar** a partir da caixa de diálogo Análise de Cluster por K-Médias.

**Máximo de Iterações.** Limita o número de iterações no algoritmo  $k$ -médias. A iteração parará após este número de iterações, mesmo se o critério de convergência não estiver satisfeito. Este número deve ser entre 1 e 999.

Para reproduzir o algoritmo utilizado pelo comando Cluster Rápido anterior à versão 5.0, configure **Máximo de Iterações** para 1.

**Critério de Convergência.** Determina quando iteração é interrompida. Ele representa uma proporção da distância mínima entre os centros de cluster iniciais, por isso deve ser maior que 0, mas não superior a 1. Se o critério for igual a 0,02, por exemplo, a iteração terminará quando uma iteração completa não mover nenhum dos centros de cluster por uma distância de mais de 2% da menor distância entre quaisquer centros de cluster iniciais.

**Utilizar médias de execução.** Permite solicitar que os centros do cluster sejam atualizados após cada caso ser designado. Se você não selecionar essa opção, novos centros do cluster serão calculados após todos os casos terem sido designados.

## Salvar Análise de Cluster por K-Médias

É possível salvar informações sobre a solução como novas variáveis para serem usadas em análises subsequentes:

**Associação de cluster.** Cria uma nova variável indicando a associação final do cluster de cada caso. Os valores da nova variável variam de 1 ao número de clusters.

**Distância do centro do cluster.** Cria uma nova variável indicando a distância euclidiana entre cada caso e seu centro de classificação.

## Opções de análise de cluster por k-médias

**Estatísticas.** É possível selecionar as seguintes estatísticas: centros de cluster iniciais, tabela ANOVA e informações de cluster para cada caso.

- *Centros de clusters iniciais.* Primeira estimativa das médias da variável para cada um dos clusters... Por padrão, um número de casos bem espaçados igual ao número de clusters é selecionado a partir dos dados. Os centros do cluster iniciais são utilizados para uma primeira rodada de classificação e, em seguida, são atualizados.
- *Tabela de ANOVA.* Exibe uma tabela de análise de variação que inclui testes F univariados para cada variável de armazenamento em cluster. Os testes F são apenas descritivos e as probabilidades resultantes não devem ser interpretadas. A tabela ANOVA não será exibida se todos os casos forem designados para um único cluster.
- *Informações de cluster para cada caso.* Exibe para cada caso a designação de cluster final e a distância euclidiana entre o caso e o centro do cluster usado para classificar o caso.. Exibe também a distância euclidiana entre os centros do cluster final.

**Valores omissos.** As opções disponíveis são **Excluir listwise dos casos** ou **Excluir Casos Entre Pares**.

- **Excluir listwise dos casos.** Exclui da análise os casos com valores ausentes para qualquer variável de clusterização.
- **Excluir casos entre pares.** Atribui casos a clusters com base em distâncias que são calculadas de todas as variáveis com valores não omissos.

## Recursos adicionais do comando QUICK CLUSTER

O procedimento Cluster de K-Médias utiliza a sintaxe de comando QUICK CLUSTER. O idioma da sintaxe de comando também permite:

- Aceitar os primeiros  $k$  casos como os centros do cluster iniciais, evitando, assim, a passagem de dados que é normalmente utilizada para estimá-los.
- Especifique os centros de cluster iniciais diretamente como parte da sintaxe do comando.
- Especifique nomes para variáveis salvas.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Testes não paramétricos

---

Testes não paramétricos fazem suposições mínimas sobre a distribuição subjacente dos dados. Os testes que estão disponíveis nesses diálogos podem ser agrupados em três amplas categorias com base em como os dados são organizados:

- Um teste de uma amostra analisa um campo.
- Um teste para amostras relacionadas compara dois ou mais campos para o mesmo conjunto de casos.
- Um teste de amostras independentes analisa um campo que é agrupado por categorias de um outro campo.

### Testes não paramétricos de amostra única

Testes não paramétricos de uma amostra identificam diferenças em campos únicos usando um ou mais testes não paramétricos. Testes não paramétricos não assumem que seus dados seguem a distribuição normal.

**Qual é o seu objetivo?** Os objetivos permitem que você especifique rapidamente configurações de teste diferentes, mas comumente usadas.

- **Comparar automaticamente dados observados com hipotéticos.** Este objetivo aplica o teste Binomial para campos categóricos com apenas duas categorias, o teste Qui-Quadrado para todos os outros campos categóricos e o teste de Kolmogorov-Smirnov para campos contínuos.
- **Sequência de teste para aleatoriedade.** Este objetivo usa o teste de sequência para testar a aleatoriedade da sequência observada de valores de dados.
- **Análise customizada.** Quando desejar alterar manualmente as configurações de teste na guia Configurações, selecione esta opção. Observe que essa configuração é selecionada automaticamente se você fizer alterações subsequentes nas opções da guia Configurações que são incompatíveis com o objetivo selecionado no momento.

### Obtendo testes não paramétricos de uma amostra

No menu, escolha:

**Analisar > Testes Não Paramétricos > Uma Amostra...**

1. Clique em **Executar**.

Opcionalmente, é possível:

- Especificar um objetivo na guia Objetivo.
- Especificar designações de campo na guia Campos.
- Especificar configurações especializadas na guia Configurações.

### Guia Campos

A guia Campos especifica quais campos devem ser testados.

**Usar papéis predefinidos.** Esta opção usa informações de campo existentes. Todos os campos com uma função predefinida como Entrada, Destino ou Ambos serão usados como campos de teste. Pelo menos um campo de teste é necessário para obter mais informações.

**Usar designações de campo customizado.** Essa opção permite substituir funções de campo. Após selecionar esta opção, especifique os campos abaixo:

- **Campos de Teste.** Selecione um ou mais campos.

### Guia Configurações

A guia Configurações abrange vários grupos diferentes de configurações que podem ser modificadas para ajustar com precisão como o algoritmo processa os seus dados. Se você fizer qualquer mudança nas

configurações padrão que for incompatível com os objetivos selecionados atualmente, a guia Objetivo será atualizada automaticamente para selecionar a opção **Customizar análise**.

## **Escolher Testes**

Essas configurações especificam os testes a serem executados nos campos especificados na guia Campos.

**Escolher automaticamente os testes com base nos dados.** Essa configuração aplica o teste Binomial a campos categóricos com apenas duas categorias válidas (não ausentes), o teste qui-quadrado a todos os outros campos categóricos e o teste de Kolmogorov-Smirnov a campos contínuos.

**Customizar testes.** Essa configuração permite que você escolha testes específicos a serem realizados.

- **Comparar probabilidade binária observada com hipotética (Teste binomial).** O Teste binomial pode ser aplicado a todos os campos. Isso produz um teste de uma amostra que testa se a distribuição observada de um campo de sinalização (um campo categórico com apenas duas categorias) é a mesma que se espera de uma distribuição binomial especificada. Além disso, é possível solicitar intervalos de confiança. Consulte [“Opções de teste binomial”](#) na página 199 para obter detalhes sobre as configurações de teste.
- **Comparar probabilidades observadas com hipotéticas (Teste Qui-Quadrado).** O teste Qui-Quadrado é aplicado a campos nominais e ordinais. Isso produz um teste de uma amostra que calcula uma estatística qui-quadrado com base nas diferenças entre as frequências observadas e esperadas das categorias de um campo. Consulte [“Opções de Testes Qui-Quadrado”](#) na página 200 para obter detalhes sobre as configurações de teste.
- **Testar distribuição observada com relação à hipotética (Teste de Kolmogorov-Smirnov).** O teste de Kolmogorov-Smirnov é aplicado a campos contínuos e ordinais. Isso produz um teste de uma amostra para saber se a função de distribuição cumulativa de amostra para um campo é homogênea com uma distribuição uniforme, normal, Poisson ou exponencial. Consulte [“Opções de Kolmogorov-Smirnov”](#) na página 200 para obter detalhes sobre as configurações de teste.
- **Comparar mediana com hipotética (Teste dos postos sinalizados de Wilcoxon).** O teste dos postos sinalizados de Wilcoxon é aplicado a campos contínuos e ordinais. Isso produz um teste de uma amostra de valor médio de um campo. Especifique um número como a mediana hipotética.
- **Sequência de teste para aleatoriedade (Teste de execuções).** O Teste de execuções é aplicado para todos os campos. Isso produz um teste de uma amostra para saber se a sequência de valores de um campo dicotomizado é aleatória. Consulte [“Opções de Teste de Execuções”](#) na página 200 para obter detalhes sobre as configurações de teste.

### *Opções de teste binomial*

O teste binomial destina-se a campos sinalizadores (campos categóricos com apenas duas categorias), mas é aplicado a todos os campos usando regras para definir "sucesso".

**Proporção hipotética.** Isso especifica a proporção esperada de registros definidos como "sucessos" ou  $p$ . Especifique um valor maior que 0 e menor que 1. O padrão é 0,5.

**Intervalo de Confiança.** Os métodos a seguir para calcular intervalos de confiança para dados binários estão disponíveis:

- **Clopper-Pearson (exato).** Um intervalo exato baseado na distribuição binomial cumulativa.
- **Jeffreys.** Um intervalo Bayesiano baseado na distribuição posterior de  $p$  usando as informações a priori de Jeffreys.
- **Razão de verossimilhança.** Um intervalo baseado na função de probabilidade para  $p$ .

**Definir Sucesso para Campos Categóricos.** Isso especifica como o "sucesso" (os valores de dados testados em relação à proporção hipotética) é definido para campos categóricos.

- **Usar a primeira categoria localizada em dados** executa o teste binomial usando o primeiro valor localizado na amostra para definir "sucesso". Esta opção só se aplica a campos nominais ou ordinais com apenas dois valores; todos os outros campos categóricos especificados na guia Campos onde esta opção é usada não serão testados. Este é o padrão.

- **Especificar valores de sucesso** executa o teste binomial usando a lista especificada de valores para definir "sucesso". Especifique uma lista de sequências de caracteres ou de valores numéricos. Os valores na lista não precisam estar presentes na amostra.

**Definir Sucesso para Campos Contínuos.** Isso especifica como o "sucesso" (os valores de dados testados em relação à proporção hipotética) é definido para campos contínuos. O sucesso é definido como valores iguais ou inferiores a um ponto de corte.

- **Ponto médio da amostra** configura o ponto de corte na média dos valores mínimo e máximo.
- **Ponto de corte customizado** permite que você especifique um valor para o ponto de corte.

#### *Opções de Testes Qui-Quadrado*

**Todas as categorias têm probabilidade igual.** Isso produz frequências iguais entre todas as categorias na amostra. Este é o padrão.

**Customizar probabilidade esperada.** Isso permite especificar frequências desiguais para uma lista específica de categorias. Especifique uma lista de sequências de caracteres ou de valores numéricos. Os valores na lista não precisam estar presentes na amostra. Na coluna **Categoria**, especifique valores de categoria. Na coluna **Frequência Parente**, especifique um valor maior que 0 para cada categoria. As frequências customizadas são tratadas como proporções para que, por exemplo, especificar as frequências 1, 2 e 3 seja equivalente a especificar as frequências 10, 20 e 30, e ambas especificam que 1/6 dos registros devem se enquadrar na primeira categoria, 1/3 no segundo e 1/2 no terceiro. Quando as probabilidades esperadas personalizadas são especificadas, os valores da categoria personalizada devem incluir todos os valores de campo nos dados; caso contrário, o teste não é executado para esse campo.

#### *Opções de Kolmogorov-Smirnov*

Esta caixa de diálogo especifica quais distribuições devem ser testadas e os parâmetros das distribuições hipotéticas.

Quando certos parâmetros da distribuição precisam ser estimados a partir da amostra, o teste de Kolmogorov-Smirnov não se aplica mais. Nestas situações, a estatística do teste de Lilliefors pode ser usada para estimar o valor  $p$  usando a amostragem de Monte Carlo para normalidade de teste com uma média e variância desconhecidas. O teste de Lilliefors se aplica às três distribuições contínuas (**Normal**, **Exponencial** e **Uniforme**). Observe que o teste não se aplicará se a distribuição subjacente for discreta (**Poisson**). O teste é definido somente para inferência de uma amostra quando os parâmetros de distribuição correspondentes não são especificados.

#### **Normal**

**Usar dados de amostra** usa a média observada e o desvio padrão e fornece opções para selecionar os resultados existentes de **Teste assintótico** ou usar **Teste de Lilliefors baseado na amostragem de Monte Carlo**. **Customizado** permite especificar valores.

#### **Uniforme**

**Usar dados de amostra** usa o mínimo e o máximo observados e usa o teste de Lilliefors baseado na amostragem de Monte Carlo. **Customizado** permite especificar valores mínimos e máximos.

#### **Exponencial**

**Média da amostra** usa a média observada e usa o teste de Lilliefors baseado na amostragem de Monte Carlo. **Customizado** permite especificar um valor médio observado.

#### **Poisson**

**Média** permite especificar um valor médio observado.

#### *Opções de Teste de Execuções*

O teste de execuções destina-se a campos sinalizadores (campos categóricos com apenas duas categorias), mas pode ser aplicado a todos os campos usando regras para definir os grupos.

**Definir Grupos para Campos Categóricos.** As seguintes opções estão disponíveis:

- **Existem somente 2 categorias na amostra** executa o teste de execuções usando os valores localizados na amostra para definir os grupos. Esta opção só se aplica a campos nominais ou ordinais com apenas



dois valores; todos os outros campos categóricos especificados na guia Campos onde esta opção é usada não serão testados.

- **Recodificar dados em 2 categorias** executa o teste de execuções usando a lista especificada de valores para definir um dos grupos. Todos os outros valores na amostra definem o outro grupo. Nem todos os valores da lista precisam estar presentes na amostra, mas pelo menos um registro deve estar em cada grupo.

**Definir Ponto de Corte para Campos Contínuos.** Isso especifica como os grupos são definidos para campos contínuos. O primeiro grupo é definido como valores iguais ou inferiores a um ponto de corte.

- **Mediana de amostra** configura o ponto de corte na mediana de amostra.
- **Média de amostra** configura o ponto de corte na média de amostra.
- **Customizado** permite que você especifique um valor para o ponto de corte.

## **Opções de Teste**

### **Nível de significância**

Isso especifica o nível de significância (alpha) para todos os testes. Especifique um valor numérico entre 0 e 1. 0,05 é o padrão.

### **Intervalo de confiança (%)**

Isso especifica o nível de confiança para todos os intervalos de confiança produzidos. Especifique um valor numérico entre 0 e 100. 95 é o padrão.

### **Casos excluídos**

Isso especifica como determinar a base de caso para testes.

#### **Excluir teste de casos por teste**

Os registros com valores ausentes para um campo usado para um teste específico são omitidos desse teste. Quando vários testes são especificados na análise, cada teste é avaliado separadamente.

#### **Excluir casos de listwise**

Os registros com valores ausentes para qualquer campo nomeado na guia Campos são excluídos de todas as análises.

## **Amostragem de Monte Carlo**

Quando certos parâmetros da distribuição precisam ser estimados a partir da amostra, o teste de Kolmogorov-Smirnov não se aplica mais. Nestas situações, a estatística do teste de Lilliefors pode ser usada para estimar o valor  $p$  usando a amostragem de Monte Carlo para normalidade de teste com uma média e variância desconhecidas. O teste de Lilliefors se aplica às três distribuições contínuas (**Normal, Exponencial e Uniforme**). Observe que o teste não se aplicará se a distribuição subjacente for discreta (**Poisson**). O teste é definido somente para inferência de uma amostra quando os parâmetros de distribuição correspondentes não são especificados.

### **Configure o valor inicial customizado**

Quando ativada, esta configuração fornece a opção de reconfigurar o valor de **Semente** aleatória que é usado para a amostragem de Monte Carlo. O valor deve ser um único inteiro entre 1 e 2.147.483.647 (o valor padrão é 2.000.000). Por padrão, a configuração não está ativada, o que produz um valor de semente aleatório.

### **Número de amostras**

Redefine o número de réplicas de amostragem de Monte Carlo que são usadas pelo teste Lilliefors. O valor deve ser um número inteiro único entre 100 e o maior número inteiro. O valor padrão é 10.000.

### **Nível de confiança de simulação (%)**

Reconfigura o nível de intervalo de confiança estimado do teste de Kolmogorov-Smirnov. O valor deve ser um valor único entre 0 e 100. O valor padrão é 99.

## Valores Omissos de Usuário

**Valores Omissos de Usuário para Campos Categóricos.** Os campos categóricos devem ter valores válidos para que um registro seja incluído na análise. Esses controles permitem que você decida se os valores omissos do usuário são tratados como válidos entre os campos categóricos. Valores ausentes do sistema e valores omissos para campos contínuos são sempre tratados como inválidos.

## Recursos adicionais do comando NPTESTS

O idioma da sintaxe de comando também permite:

- Especificar testes de uma amostra, de amostras independentes e de amostras relacionadas em uma única execução do procedimento.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Testes Não Paramétricos de Amostras Independentes

Testes não paramétricos de amostras independentes identificam diferenças entre dois ou mais grupos usando um ou mais testes não paramétricos. Testes não paramétricos não assumem que seus dados seguem a distribuição normal.

**Qual é o seu objetivo?** Os objetivos permitem que você especifique rapidamente configurações de teste diferentes, mas comumente usadas.

- **Comparar automaticamente distribuições ao longo de grupos.** Esse objetivo aplica-se ao teste U de Mann-Whitney para dados com 2 grupos ou ao ANOVA de 1 via de Kruskal-Wallis para dados com grupos  $k$ .
- **Comparar medianas ao longo de grupos.** Este objetivo usa o teste da mediana para comparar as medianas observadas entre os grupos.
- **Análise customizada.** Quando desejar alterar manualmente as configurações de teste na guia Configurações, selecione esta opção. Observe que essa configuração é selecionada automaticamente se você fizer alterações subsequentes nas opções da guia Configurações que são incompatíveis com o objetivo selecionado no momento.

## Para Obter Testes Não Paramétricos de Amostras Independentes

A partir dos menus, escolha:

**Analisar > Testes Não Paramétricos > Amostras Independentes...**

1. Clique em **Executar**.

Opcionalmente, é possível:

- Especificar um objetivo na guia Objetivo.
- Especificar designações de campo na guia Campos.
- Especificar configurações especialistas na guia Configurações.

## Guia Campos

A guia Campos especifica quais campos devem ser testados e o campo usado para definir grupos.

**Usar papéis predefinidos.** Esta opção usa informações de campo existentes. Todos os campos contínuos e ordinais com uma função predefinida como Resposta ou Ambos serão usados como campos de teste. Se houver um único campo categórico com uma função predefinida como Entrada, ele será usado como campo de agrupamento. Caso contrário, nenhum campo de agrupamento é usado por padrão e você deve usar atribuições de campo customizado. Pelo menos um campo de teste e um campo de agrupamento são obrigatórios. para obter mais informações.

**Usar designações de campo customizado.** Essa opção permite substituir funções de campo. Após selecionar esta opção, especifique os campos abaixo:

- **Campos de Teste.** Selecione um ou mais campos contínuos ou ordinais.
- **Grupos.** Selecione um campo categórico.

## Guia Configurações

A guia Configurações abrange vários grupos diferentes de configurações que podem ser modificadas para ajustar com precisão como o algoritmo processa os seus dados. Se você fizer qualquer mudança nas configurações padrão que for incompatível com os objetivos selecionados atualmente, a guia Objetivo será atualizada automaticamente para selecionar a opção **Customizar análise**.

### Escolher testes

Essas configurações especificam os testes a serem executados nos campos especificados na guia Campos.

**Escolher automaticamente os testes com base nos dados.** Essa configuração aplica-se ao teste U de Mann-Whitney para dados com 2 grupos ou ao ANOVA de 1 via de Kruskal-Wallis para dados com grupos  $k$ .

**Customizar testes.** Essa configuração permite que você escolha testes específicos a serem realizados.

- **Comparar Distribuições ao longo de Grupos.** Estes produzem testes de amostras independentes para saber se as amostras são da mesma população.

**Mann-Whitney U (2 amostras)** usa o ranqueamento de cada caso para testar se os grupos são desenhados a partir da mesma população. O primeiro valor em ordem crescente do campo de agrupamento define o primeiro grupo e o segundo define o segundo grupo. Se o campo de agrupamento tiver mais de dois valores, esse teste não será produzido.

**Kolmogorov-Smirnov (2 amostras)** é sensível a qualquer diferença em mediana, dispersão, assimetria e assim por diante, entre as duas distribuições. Se o campo de agrupamento tiver mais de dois valores, esse teste não será produzido.

**Sequência de teste para aleatoriedade (Wald-Wolfowitz para 2 amostras)** produz um teste de execuções com associação ao grupo como o critério. Se o campo de agrupamento tiver mais de dois valores, esse teste não será produzido.

**ANOVA de 1 via de Kruskal-Wallis (amostras  $k$ )** é uma extensão do teste U de Mann-Whitney e o analógico não paramétrico de análise de variância para um fator. É possível opcionalmente solicitar várias comparações das amostras  $k$ , todas as várias comparações **entre pares** ou comparações **de redução de stepwise**.

**Teste para alternativas ordenadas (Jonckheere-Terpstra para amostras  $k$ )** é uma alternativa mais eficaz para Kruskal-Wallis quando as amostras  $k$  têm uma ordem natural. Por exemplo, as populações  $k$  podem representar  $k$  aumentando as temperaturas. A hipótese de que temperaturas diferentes produzem a mesma distribuição de resposta é testada contra a alternativa de que à medida que a temperatura aumenta, a magnitude da resposta aumenta. Aqui, a hipótese alternativa é ordenada; portanto, Jonckheere-Terpstra é o teste mais apropriado para uso. **Menor para maior** especifica a hipótese alternativa de que o parâmetro de localização do primeiro grupo é menor que ou igual ao segundo, que é menor que ou igual ao terceiro, e assim por diante. **Maior para menor** especifica a hipótese alternativa de que o parâmetro de localização do primeiro grupo é maior que ou igual ao segundo, que é maior que ou igual ao terceiro, e assim por diante. Para ambas as opções, a hipótese alternativa também assume que as localizações não são todas iguais. É possível opcionalmente solicitar várias comparações das amostras  $k$ , todas as várias comparações **entre pares** ou comparações **de redução de stepwise**.

- **Comparar Intervalos ao longo de Grupos.** Isso produz testes de amostras independentes para verificar se as amostras têm o mesmo intervalo. **Reação extrema de Moisés (2 amostras)** testa um grupo de controle versus um grupo de comparação. O primeiro valor em ordem crescente do campo de agrupamento define o grupo de controle e o segundo define o grupo de comparação. Se o campo de agrupamento tiver mais de dois valores, esse teste não será produzido.

- **Comparar Medianas ao longo de Grupos.** Isso produz testes de amostras independentes para saber se as amostras têm a mesma mediana. **Teste de mediana (amostras k)** podem usar a mediana de amostra agrupada (calculada ao longo de todos os registros no conjunto de dados) ou um valor customizado como a mediana hipotética. É possível opcionalmente solicitar várias comparações das amostras *k*, todas as várias comparações **entre pares** ou comparações **de redução de stepwise**.
- **Estimar Intervalos de Confiança ao longo de Grupos.** A **estimativa de Hodges-Lehman (2 amostras)** produz uma estimativa de amostras independentes e um intervalo de confiança para a diferença nas medianas de dois grupos. Se o campo de agrupamento tiver mais de dois valores, esse teste não será produzido.

### Opções de testes

**Nível de significância.** Isso especifica o nível de significância (alpha) para todos os testes. Especifique um valor numérico entre 0 e 1. 0,05 é o padrão.

**Intervalo de confiança (%).** Isso especifica o nível de confiança para todos os intervalos de confiança produzidos. Especifique um valor numérico entre 0 e 100. 95 é o padrão.

**Casos Excluídos.** Isso especifica como determinar a base de caso para testes. **Excluir casos listwise** significa que registros com valores omissos para qualquer campo que é nomeado em qualquer subcomando são excluídos de todas as análises. **Excluir teste de casos por teste** significa que registros com valores omissos para um campo que é usado para um teste específico são omitidos desse teste. Quando vários testes são especificados na análise, cada teste é avaliado separadamente.

### Valores Omissos de Usuário

**Valores Omissos de Usuário para Campos Categóricos.** Os campos categóricos devem ter valores válidos para que um registro seja incluído na análise. Esses controles permitem que você decida se os valores omissos do usuário são tratados como válidos entre os campos categóricos. Valores omissos do sistema e valores omissos para campos contínuos são sempre tratados como inválidos.

### Recursos adicionais do comando NPTESTS

O idioma da sintaxe de comando também permite:

- Especificar testes de uma amostra, de amostras independentes e de amostras relacionadas em uma única execução do procedimento.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Testes não paramétricos de amostras relacionadas

Identifica diferenças entre dois ou mais campos relacionados usando um ou mais testes não paramétricos. Testes não paramétricos não assumem que seus dados seguem a distribuição normal.

**Considerações de Dados.** Cada registro corresponde a um determinado assunto para o qual duas ou mais medidas relacionadas estão armazenadas em campos separados no conjunto de dados. Por exemplo, um estudo referente à efetividade de um plano de dieta pode ser analisado usando testes não paramétricos de amostras relacionadas se cada ponderação do assunto for medida em intervalos regulares e armazenada em campos como *Peso pré-dieta*, *Peso intermediário* e *Peso pós-dieta*. Esses campos são "relacionados".

**Qual é o seu objetivo?** Os objetivos permitem que você especifique rapidamente configurações de teste diferentes, mas comumente usadas.

- **Comparar automaticamente dados observados com dados hipotéticos.** Este objetivo aplica o teste de McNemar para dados categóricos quando 2 campos são especificados, o Q de Cochran para dados categóricos quando mais de 2 campos são especificados, o teste Wilcoxon Matched-Pair Signed-Rank para dados contínuos quando 2 campos são especificados e o teste 2-Way de Friedman ANOVA por Classificações para dados contínuos quando mais de 2 campos são especificados.
- **Análise customizada.** Quando desejar alterar manualmente as configurações de teste na guia Configurações, selecione esta opção. Observe que essa configuração é selecionada automaticamente

se você fizer alterações subsequentes nas opções da guia Configurações que são incompatíveis com o objetivo selecionado no momento.

Quando campos de diferentes níveis de medição são especificados, eles são separados primeiro por nível de medição e, em seguida, o teste apropriado é aplicado a cada grupo. Por exemplo, se você escolher **Comparar automaticamente dados observados com dados hipotéticos** com o seu objetivo e especificar 3 campos contínuos e 2 campos nominais, então o teste de Friedman é aplicado nos campos contínuos e o teste de McNemar é aplicado nos campos nominais.

## Para Obter Testes Não Paramétricos de Amostras Relacionadas

A partir dos menus, escolha:

**Analisar > Testes Não Paramétricos > Amostras Relacionadas...**

1. Clique em **Executar**.

Opcionalmente, é possível:

- Especificar um objetivo na guia Objetivo.
- Especificar designações de campo na guia Campos.
- Especificar configurações especialistas na guia Configurações.

## Guia Campos

A guia Campos especifica quais campos devem ser testados.

**Usar papéis predefinidos.** Esta opção usa informações de campo existentes. Todos os campos com uma função predefinida como Destino ou Ambos serão usados como campos de teste. Pelo menos dois campos de teste são necessários. para obter mais informações.

**Usar designações de campo customizado.** Essa opção permite substituir funções de campo. Após selecionar esta opção, especifique os campos abaixo:

- **Campos de Teste.** Selecione dois ou mais campos. Cada campo corresponde a uma amostra relacionada separada.

## Guia Configurações

A guia Configurações compreende vários grupos diferentes de configurações que você pode modificar para ajustar como o procedimento processa seus dados. Se você fizer qualquer mudança nas configurações padrão que for incompatível com os outros objetivos, a guia Objetivo será atualizada automaticamente para selecionar a opção **Customizar análise**.

### **Escolher Testes**

Essas configurações especificam os testes a serem executados nos campos especificados na guia Campos.

**Escolher automaticamente os testes com base nos dados.** Esta configuração aplica o teste de McNemar a dados categóricos quando 2 campos são especificados, o Q de Cochran a dados categóricos quando mais de 2 campos são especificados, o teste Wilcoxon Matched-Pair Signed-Rank para dados contínuos quando 2 campos são especificados e o teste de 2 vias de Friedman ANOVA por Classificações para dados contínuos quando mais de 2 campos são especificados.

**Customizar testes.** Essa configuração permite que você escolha testes específicos a serem realizados.

- **Teste para mudança em dados binários. O teste de McNemar (2 amostras)** pode ser aplicado a campos categóricos. Isso produz um teste de amostras relacionadas para saber se as combinações de valores entre dois campos de sinalizadores (campos categóricos com apenas dois valores) são igualmente prováveis. Se houver mais de dois campos especificados na guia Campos, esse teste não será executado. Consulte “Teste de McNemar: Definir Sucesso” na página 206 para obter detalhes sobre as configurações de teste. **Q de Cochran (amostras k)** pode ser aplicado em campos categóricos.

Isso produz um teste de amostras relacionadas de se combinações de valores entre campos de sinalização  $k$  (campos categóricos com apenas dois valores) são igualmente prováveis. É possível opcionalmente solicitar várias comparações das amostras  $k$ , todas as várias comparações **entre pares** ou comparações **de redução de stepwise**. Consulte “Q de Cochran: Definir sucesso” na página 206 para obter detalhes sobre as configurações de teste.

- **Teste para Mudanças em Dados Multinomiais. O teste de homogeneidade marginal (2 amostras)** produz um teste de amostras relacionadas para verificar se as combinações de valores entre dois campos ordinais emparelhados são igualmente prováveis. O teste de homogeneidade marginal é normalmente usado em situações de medidas repetidas. Este teste é uma extensão do teste de McNemar de resposta binária para resposta multinomial. Se houver mais de dois campos especificados na guia Campos, esse teste não será executado.
- **Comparar Diferença de Mediana com Hipotética.** Cada um desses testes produz um teste de amostras relacionadas para saber se a diferença mediana entre dois é diferente de 0. O teste aplica-se a campos contínuos e ordinais. Se houver mais do que dois campos especificados na guia Campos, esses testes não serão realizados.
- **Estimar Intervalos de Confiança.** Isso produz uma estimativa de amostras relacionadas e um intervalo de confiança para a diferença de mediana entre dois campos emparelhados. O teste aplica-se a campos contínuos e ordinais. Se houver mais do que dois campos especificados na guia Campos, esse teste não será realizado.
- **Quantificar Associações. Kendall's coefficient of concordance (k samples)** produz uma medida de acordo entre julgadores ou avaliadores, onde cada registro é uma avaliação de vários itens (campos). É possível opcionalmente solicitar várias comparações das amostras  $k$ , todas as várias comparações **entre pares** ou comparações **de redução de stepwise**.
- **Comparar distribuições. A ANOVA de 2 vias de Friedman por ranks (amostras k)** produz um teste de amostras relacionadas para saber se amostras  $k$  relacionadas foram extraídas da mesma população. É possível opcionalmente solicitar várias comparações das amostras  $k$ , todas as várias comparações **entre pares** ou comparações **de redução de stepwise**.

#### *Teste de McNemar: Definir Sucesso*

O teste de McNemar destina-se a campos de sinalização (campos categóricos com apenas duas categorias), mas é aplicado a todos os campos categóricos usando regras para definir "sucesso".

**Definir Sucesso para Campos Categóricos.** Isso especifica como "sucesso" é definido para campos categóricos.

- **Usar a primeira categoria localizada em dados** executa o teste usando o primeiro valor localizado na amostra para definir "sucesso". Esta opção só se aplica a campos nominais ou ordinais com apenas dois valores; todos os outros campos categóricos especificados na guia Campos onde esta opção é usada não serão testados. Este é o padrão.
- **Especificar valores de sucesso** executa o teste usando a lista especificada de valores para definir "sucesso". Especifique uma lista de sequências de caracteres ou de valores numéricos. Os valores na lista não precisam estar presentes na amostra.

#### *Q de Cochran: Definir sucesso*

O teste Q de Cochran destina-se a campos sinalizadores (campos categóricos com apenas duas categorias), mas é aplicado a todos os campos categóricos usando regras para definir "sucesso".

**Definir Sucesso para Campos Categóricos.** Isso especifica como "sucesso" é definido para campos categóricos.

- **Usar a primeira categoria localizada em dados** executa o teste usando o primeiro valor localizado na amostra para definir "sucesso". Esta opção só se aplica a campos nominais ou ordinais com apenas dois valores; todos os outros campos categóricos especificados na guia Campos onde esta opção é usada não serão testados. Este é o padrão.
- **Especificar valores de sucesso** executa o teste usando a lista especificada de valores para definir "sucesso". Especifique uma lista de sequências de caracteres ou de valores numéricos. Os valores na lista não precisam estar presentes na amostra.

## Opções de Teste

**Nível de significância.** Isso especifica o nível de significância (alpha) para todos os testes. Especifique um valor numérico entre 0 e 1. 0,05 é o valor padrão.

**Intervalo de confiança (%).** Isso especifica o nível de confiança para todos os intervalos de confiança produzidos. Especifique um valor numérico entre 0 e 100. 95 é o valor padrão.

**Casos Excluídos.** Isso especifica como determinar a base de caso para testes.

- **Excluir casos listwise** significa que registros com valores omissos para qualquer campo que é nomeado em qualquer subcomando são excluídos de todas as análises.
- **Excluir teste de casos por teste** significa que registros com valores omissos para um campo que é usado para um teste específico são omitidos desse teste. Quando vários testes são especificados na análise, cada teste é avaliado separadamente.

## Valores Omissos de Usuário

**Valores Omissos de Usuário para Campos Categóricos.** Os campos categóricos devem ter valores válidos para que um registro seja incluído na análise. Esses controles permitem que você decida se os valores omissos do usuário são tratados como válidos entre os campos categóricos. Valores omissos do sistema e valores omissos para campos contínuos são sempre tratados como inválidos.

## Recursos adicionais do comando NPTESTS

O idioma da sintaxe de comando também permite:

- Especificar testes de uma amostra, de amostras independentes e de amostras relacionadas em uma única execução do procedimento.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Recursos adicionais do comando NPTESTS

O idioma da sintaxe de comando também permite:

- Especificar testes de uma amostra, de amostras independentes e de amostras relacionadas em uma única execução do procedimento.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Diálogos Legado

Há uma série de diálogos "anteriores" que também executam testes não paramétricos. Esses diálogos suportam a funcionalidade fornecida pela opção de Testes Exatos.

**Teste Qui-Quadrado.** Tabula uma variável em categorias e calcula uma estatística qui-quadrado baseada nas diferenças entre as frequências observadas e esperadas.

**Teste Binomial.** Compara a frequência observada em cada categoria de uma variável dicotômica com as frequências esperadas da distribuição binomial.

**Teste de Execuções.** Testa se a ordem de ocorrência de dois valores de uma variável é aleatória.

**Teste de Kolmogorov-Smirnov de Uma Amostra.** Compara a função de distribuição cumulativa observada para uma variável com uma distribuição teórica especificada, que pode ser normal, uniforme, exponencial ou Poisson.

**Testes de Duas Amostras Independentes.** Compara dois grupos de casos em uma variável. O teste  $U$  de Mann-Whitney, o teste de Kolmogorov-Smirnov de duas amostras, o teste de Moisés de reações extremas e o teste de sequências de Wald-Wolfowitz estão disponíveis.

**Testes de Duas Amostras Relacionadas.** Compara as distribuições de duas variáveis. O teste de classificação assinada de Wilcoxon, o teste de assinatura e o teste de McNemar estão disponíveis.

**Testes para Várias Amostras Independentes.** Compara dois ou mais grupos de casos em uma variável. O teste de Kruskal-Wallis, o teste Median e o teste Jonckheere-Terpstra estão disponíveis.

**Testes para Várias Amostras Relacionadas.** Compara as distribuições de duas ou mais variáveis. O teste de Friedman, *W* de Kendall e *Q* de Cochran estão disponíveis.

Os quartis e a média, desvio padrão, mínimo, máximo e número de casos não omissos estão disponíveis para todos os testes acima.

## Teste qui-quadrado

O procedimento de teste de qui-quadrado tabula uma variável em categorias e calcula uma estatística de qui-quadrado. Este teste de adequação compara as frequências observadas e esperadas em cada categoria para testar se todas as categorias contêm a mesma proporção de valores ou testar que cada categoria contém uma proporção de valores especificada pelo usuário.

**Exemplos.** O teste do qui-quadrado pode ser usado para determinar se um saco de jujubas contém proporções iguais de balas azuis, marrons, verdes, laranja, vermelhas e amarelas. Também é possível testar para ver se um saco de jujubas contém 5% de doces azuis, 30% marrons, 10% verdes, 20% laranja, 15% vermelhos e 15% amarelos.

**Estatísticas.** Média, desvio padrão, mínimo, máximo e quartis. O número e a porcentagem de casos não ausentes e ausentes; o número de casos observados e esperados para cada categoria; resíduos; e a estatística qui-quadrado.

Considerações de Dados de Teste Qui-Quadrado

**Dados.** Use variáveis categóricas numéricas ordenadas ou não ordenadas (níveis ordinais ou nominais de medição). Para converter variáveis de string em variáveis numéricas, use o procedimento Automatic Recode, que está disponível no menu Transform.

**Suposições.** Testes não paramétricos não requerem suposições sobre a forma da distribuição subjacente. Os dados são considerados como uma amostra aleatória. As frequências esperadas para cada categoria devem ser de pelo menos 1. Não mais de 20% das categorias devem ter frequências esperadas inferiores a 5.

Para obter um teste de qui-quadrado

1. A partir dos menus, escolha:

**Analisar > Testes Não Paramétricos > Diálogos Anteriores > Qui-Quadrado...**

2. Selecione uma ou mais variáveis de teste. Cada variável produz um teste separado.

3. Opcionalmente, clique em **Opções** para estatísticas descritivas, quartis e controle do tratamento de dados omissos.

## Intervalo e Valores Esperados de Teste Qui-Quadrado

**Intervalo Esperado.** Por padrão, cada valor distinto da variável é definido como uma categoria. Para estabelecer categorias dentro de um intervalo específico, selecione **Usar intervalo especificado** e insira valores de número inteiro para limites inferior e superior. As categorias são estabelecidas para cada valor inteiro dentro do intervalo inclusivo e os casos com valores fora dos limites são excluídos. Por exemplo, se você especificar um valor de 1 para Inferior e um valor de 4 para Superior, somente os valores inteiros de 1 a 4 serão usados para o teste qui-quadrado.

**Valores Esperados.** Por padrão, todas as categorias têm valores esperados iguais. As categorias podem ter proporções esperadas especificadas pelo usuário. Selecione **Valores**, insira um valor que seja maior que 0 para cada categoria da variável de teste e, em seguida, clique em **Incluir**. Sempre que um valor é incluído, ele aparece no final da lista de valores. A ordem dos valores é importante; corresponde à ordem crescente dos valores de categoria da variável de teste. O primeiro valor da lista corresponde ao valor mais baixo do grupo da variável de teste e o último valor corresponde ao valor mais alto. Os elementos da lista de valores são somados e, em seguida, cada valor é dividido por essa soma para calcular a proporção de casos esperados na categoria correspondente. Por exemplo, uma lista de valores de 3, 4, 5, 4 especifica as proporções esperadas de 3/16, 4/16, 5/16 e 4/16.



## Opções de Teste Qui-Quadrado

**Estatísticas.** É possível escolher uma ou ambas as estatísticas básicas.

- **Descritivo.** Exibe a média, o desvio padrão, o mínimo, o máximo e o número de casos não desconhecidos.
- **Quartis.** Exibe os valores correspondentes aos percentis 25, 50 e 75.

**Valores omissos.** Controla o tratamento de valores ausentes.

- **Excluir casos de teste por teste.** Quando vários testes são especificados, cada teste é avaliado separadamente para valores ausentes.
- **Excluir listwise dos casos.** Casos com valores ausentes para qualquer variável são excluídos de todas as análises.

## Recursos Adicionais do Comando NPAR TESTS (Teste Qui-Quadrado)

O idioma da sintaxe de comando também permite:

- Especificar valores mínimo e máximo ou frequências esperadas diferentes para variáveis diferentes (com o subcomando CHISQUARE).
- Testar a mesma variável em diferentes frequências esperadas ou utilizar intervalos diferentes (com o subcomando EXPECTED).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Teste binomial

O procedimento do Teste binomial compara as frequências observadas das duas categorias de uma variável dicotômica com as frequências que são esperadas em uma distribuição binomial com um parâmetro de probabilidade especificado. Por padrão, o parâmetro de probabilidade para ambos os grupos é 0,5. Para mudar as probabilidades, é possível inserir uma proporção de teste para o primeiro grupo. A probabilidade para o segundo grupo será 1 menos a probabilidade especificada para o primeiro grupo.

**Exemplo.** Quando você joga uma moeda de dez centavos, a probabilidade de uma cara é igual a 1/2. Com base nessa hipótese, uma moeda é lançada 40 vezes e os resultados são registrados (cara ou coroa). A partir do teste binomial, é possível descobrir que 3/4 dos lançamentos foram caras e que o nível de significância observado é pequeno (0,0027). Esses resultados indicam que não é provável que a probabilidade de uma cara seja igual a 1/2; a moeda é provavelmente tendenciosa.

**Estatísticas.** Média, desvio padrão, mínimo, máximo, número de casos não omissos e quartis.

Considerações sobre dados de teste binomial

**Dados.** As variáveis que são testadas devem ser numéricas e dicotômicas. Para converter variáveis de string em variáveis numéricas, use o procedimento Automatic Recode, que está disponível no menu Transform. Uma **variável dicotômica** é uma variável que pode ter apenas dois valores possíveis: *sim* ou *não*, *true* ou *false*, 0 ou 1, e assim por diante. O primeiro valor encontrado no conjunto de dados define o primeiro grupo e o outro valor define o segundo grupo. Se as variáveis não forem dicotômicas, deve-se especificar um ponto de corte. O ponto de corte atribui casos com valores menores ou iguais ao ponto de corte ao primeiro grupo e atribui o restante dos casos ao segundo grupo.

**Suposições.** Testes não paramétricos não requerem suposições sobre a forma da distribuição subjacente. Os dados são considerados como uma amostra aleatória.

Para obter um teste binomial

1. A partir dos menus, escolha:

**Analisar > Testes Não Paramétricos > Diálogos Anteriores > Binomial...**

2. Selecione uma ou mais variáveis de teste numérico.

3. Opcionalmente, clique em **Opções** para estatísticas descritivas, quartis e controle do tratamento de dados omissos.

### ***Opções de teste binomial***

**Estatísticas.** É possível escolher uma ou ambas as estatísticas básicas.

- **Descritivo.** Exibe a média, desvio padrão, mínimo, máximo e número de casos não ausentes.
- **Quartis.** Exibe os valores correspondentes aos percentis 25, 50 e 75.

**Valores omissos.** Controla o tratamento de valores ausentes.

- **Excluir casos de teste por teste.** Quando vários testes são especificados, cada teste é avaliado separadamente para valores ausentes.
- **Excluir listwise dos casos.** Casos com valores ausentes para qualquer variável testada são excluídos de todas as análises.

### ***Recursos Adicionais do Comando NPAR TESTS (Teste Binomial)***

O idioma da sintaxe de comando também permite:

- Selecionar grupos específicos (e excluir outros grupos) quando uma variável tiver mais de duas categorias (com o subcomando BINOMIAL).
- Especificar diferentes pontos de corte ou probabilidades para variáveis diferentes (com o subcomando BINOMIAL).
- Testar a mesma variável em diferentes pontos de corte ou probabilidades (com o subcomando EXPECTED).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

### **Teste de Sequência**

O procedimento Teste de Execuções testa se a ordem de ocorrência de dois valores de uma variável é aleatória. Uma execução é uma sequência de observações semelhantes. Uma amostra com muitas ou poucas execuções sugere que a amostra não é aleatória.

**Exemplos.** Suponha que 20 pessoas sejam reunidas para descobrir se elas comprariam um produto. A aleatoriedade suposta da amostra seria questionada seriamente se todas as 20 pessoas fossem do mesmo sexo. O teste de execuções pode ser utilizado para determinar se a amostra foi tirada aleatoriamente.

**Estatísticas.** Média, desvio padrão, mínimo, máximo, número de casos não desconhecidos e quartis.

Considerações de dados de teste de sequência

**Dados.** As variáveis devem ser numéricas. Para converter variáveis de sequência de caracteres em variáveis numéricas, utilize o procedimento Recodificação Automática, que está disponível no menu Transformação.

**Suposições.** Testes não paramétricos não requerem suposições sobre a forma da distribuição subjacente. Use amostras de distribuições de probabilidade contínuas.

Para obter um teste de sequência

1. A partir dos menus, escolha:

**Analisar > Testes Não Paramétricos > Diálogos Anteriores > Execuções...**

2. Selecione uma ou mais variáveis de teste numéricas.
3. Opcionalmente, clique em **Opções** para estatísticas descritivas, quartis e controle do tratamento de dados omissos.

## **Executa o ponto de corte de teste**

**Ponto de corte.** Especifica um ponto de corte para dicotomizar as variáveis que você escolheu. É possível utilizar a média, a mediana ou o modo observado ou utilizar um valor especificado como um ponto de corte. Casos com valores menores que o ponto de corte são atribuídos a um grupo e casos com valores maiores ou iguais ao ponto de corte são atribuídos a outro grupo. Um teste é realizado para cada ponto de corte escolhido.

## **Opções de teste de sequência**

**Estatísticas.** É possível escolher uma ou ambas as estatísticas básicas.

- **Descritivo.** Exibe a média, desvio padrão, mínimo, máximo e número de casos não ausentes.
- **Quartis.** Exibe os valores correspondentes aos percentis 25, 50 e 75.

**Valores omissos.** Controla o tratamento de valores ausentes.

- **Excluir casos de teste por teste.** Quando vários testes são especificados, cada teste é avaliado separadamente para valores ausentes.
- **Excluir listwise dos casos.** Casos com valores ausentes para qualquer variável são excluídos de todas as análises.

## **Recursos Adicionais do Comando NPAR TESTS (Teste de Execuções)**

O idioma da sintaxe de comando também permite:

- Especificar diferentes pontos de corte para variáveis diferentes (com o subcomando RUNS).
- Testar a mesma variável em diferentes pontos de corte customizados (com o subcomando RUNS).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## **Teste de Kolmogorov-Smirnov de uma amostra**

O procedimento do teste de Kolmogorov-Smirnov de uma amostra compara a função de distribuição cumulativa observada para uma variável com uma distribuição teórica especificada, que pode ser normal, uniforme, Poisson ou exponencial. O  $Z$  de Kolmogorov-Smirnov é calculado a partir da diferença maior (em valor absoluto) entre as funções de distribuição cumulativa observadas e teóricas. Este teste de ajuste testa se as observações poderiam razoavelmente ter vindo da distribuição especificada.

A partir da versão 27.0, a estatística do teste de Lilliefors pode ser usada para estimar o valor  $p$  usando a amostragem de Monte Carlo para teste em uma distribuição normal com parâmetros estimados (esta funcionalidade era possível anteriormente apenas por meio do procedimento de Exploração).

### **Exemplo**

Muitos testes paramétricos requerem variáveis normalmente distribuídas. O teste de Kolmogorov-Smirnov de uma amostra pode ser utilizado para testar se uma variável (por exemplo, *income*) é normalmente distribuída.

### **Estatísticas**

Média, desvio padrão, mínimo, máximo, número de casos não ausentes, quartis, teste de Lilliefors e simulação de Monte Carlo.

## **Considerações sobre dados de teste de Kolmogorov-Smirnov de uma amostra**

### **Dados**

Utilize variáveis quantitativas (nível de medição de intervalo ou razão).

### **Suposições**

O teste Kolmogorov-Smirnov assume que os parâmetros da distribuição do teste são especificados antecipadamente. Este procedimento estima os parâmetros a partir da amostra. A média da amostra e o desvio padrão da amostra são os parâmetros para uma distribuição normal, os valores mínimo e máximo da amostra definem o intervalo da distribuição uniforme, a média amostral é o parâmetro

para a distribuição de Poisson e a média amostral é o parâmetro para a distribuição exponencial. O poder do teste para detectar desvios da distribuição hipotética pode ser seriamente diminuído.

Quando certos parâmetros da distribuição precisam ser estimados a partir da amostra, o teste de Kolmogorov-Smirnov não se aplica mais. Nestas situações, a estatística do teste de Lilliefors pode ser usada para estimar o valor  $p$  usando a amostragem de Monte Carlo para normalidade de teste com uma média e variância desconhecidas. O teste de Lilliefors se aplica às três distribuições contínuas (**Normal**, **Exponencial** e **Uniforme**). Observe que o teste não se aplicará se a distribuição subjacente for discreta (**Poisson**). O teste é definido somente para inferência de uma amostra quando os parâmetros de distribuição correspondentes não são especificados.

## Obtendo um teste de Kolmogorov-Smirnov de uma amostra

1. A partir dos menus, escolha:

**Analisar > Testes Não Paramétricos > Diálogos Anteriores > K-S de 1 Amostra...**

2. Selecione uma ou mais variáveis de teste numérico. Cada variável produz um teste separado.

3. Opcionalmente, selecione um método de distribuição de teste:

### Normal

Quando selecionado, é possível especificar se os parâmetros de distribuição são estimados a partir de dados de amostra (a configuração padrão) ou de configurações customizadas. Quando **Usar dados de amostra** for selecionado, os resultados assintóticos existentes e a correção de significância de Lilliefors baseados na amostragem de Monte Carlo serão usados. Quando **Customizado** for selecionado, forneça valores para **Média** e **Desvio padrão**.

### Uniforme

Quando selecionado, é possível especificar se os parâmetros de distribuição são estimados a partir de dados de amostra (a configuração padrão) ou de configurações customizadas. Quando **Usar dados de amostra** for selecionado, o teste de Lilliefors será usado. Quando **Customizado** for selecionado, forneça valores para **Mín.** e **Máx.**

### Poisson

Quando selecionada, especifique um valor de parâmetro de **Média**.

### Exponencial

Quando selecionado, é possível especificar se os parâmetros de distribuição são estimados a partir da média da amostra (a configuração padrão) ou de configurações customizadas. Quando **Usar dados de amostra** for selecionado, o teste de Lilliefors será usado. Quando **Customizado** for selecionado, forneça um valor de parâmetro de **Média**.

4. Opcionalmente, clique em **Simulação** para especificar parâmetros de simulação de Monte Carlo, clique em **Exato** para especificar parâmetros de teste exatos ou clique em **Opções** para estatísticas descritivas, quartis e controle do tratamento de dados omissos.

## Teste de Kolmogorov-Smirnov de uma amostra: Simulação

Quando certos parâmetros da distribuição precisam ser estimados a partir da amostra, o teste de Kolmogorov-Smirnov não se aplica mais. Nestas situações, a estatística do teste de Lilliefors pode ser usada para estimar o valor  $p$  usando a amostragem de Monte Carlo para normalidade de teste com uma média e variância desconhecidas. O teste de Lilliefors se aplica às três distribuições contínuas (**Normal**, **Exponencial** e **Uniforme**). Observe que o teste não se aplicará se a distribuição subjacente for discreta (**Poisson**). O teste é definido somente para inferência de uma amostra quando os parâmetros de distribuição correspondentes não são especificados.

### Parâmetros de simulação de Monte Carlo

#### Nível de confiança

Essa configuração opcional redefine o nível do intervalo de confiança que é estimado pelo teste Kolmogorov-Smirnov ao usar as simulações de Monte Carlo. O valor deve estar entre 0 e 100. A configuração padrão é 99.

### **Número de amostras**

Essa configuração opcional redefine o número de réplicas que o teste Lilliefors usa para a amostragem de Monte Carlo. O valor deve ser um único inteiro entre 10.000 e o maior número de valores de amostras. O valor-padrão é 10000.

### **Suprimir os resultados de Monte Carlo para a distribuição normal**

Essa configuração opcional suprime a amostragem de Monte Carlo para os resultados de distribuição normal. Por padrão, a configuração não é selecionada (o que significa que os resultados assintóticos existentes e os resultados do teste Lilliefors, que são baseados na amostragem de Monte Carlo, são apresentados).

## **Teste de Kolmogorov-Smirnov de uma amostra: Opções**

### **Estatísticas**

É possível escolher uma ou ambas as estatísticas básicas.

### **Descritivo**

Exibe a média, desvio padrão, mínimo, máximo e número de casos não omissos.

### **Quartis**

Exibe os valores correspondentes aos percentis 25, 50 e 75.

### **Valores omissos**

Controla o tratamento de valores ausentes.

### **Excluir casos de teste a teste**

Quando vários testes são especificados, cada teste é avaliado separadamente para valores ausentes.

### **Excluir casos de listwise**

Casos com valores ausentes para qualquer variável são excluídos de todas as análises.

## **Recursos adicionais do comando NPAR TESTS (teste de Kolmogorov-Smirnov de uma amostra)**

A linguagem de sintaxe de comando também permite especificar os parâmetros da distribuição de teste (com o subcomando K-S).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## **Testes de duas amostras independentes**

O procedimento Testes de duas amostras independentes compara dois grupos de casos em uma variável.

**Exemplo.** Novos aparelhos dentários foram desenvolvidos com o objetivo de serem mais confortáveis, parecerem melhores e proporcionarem um progresso mais rápido no realinhamento dos dentes. Para descobrir se o novo aparelho deve ser usado tanto quanto o antigo, 10 crianças são escolhidas aleatoriamente para usar o aparelho antigo e outras 10 crianças são escolhidas para usar o novo aparelho. No teste  $U$  de Mann-Whitney, você pode descobrir que, em média, as crianças com os novos aparelhos não precisaram utilizá-los tanto quanto os aparelhos antigos.

**Estatísticas.** Média, desvio padrão, mínimo, máximo, número de casos não omissos e quartis. Testes:  $U$  de Mann-Whitney, reações extremas de Moses,  $Z$  de Kolmogorov-Smirnov, execuções de Wald-Wolfowitz.

Considerações de Dados de Testes de Duas Amostras Independentes

**Dados.** Use variáveis numéricas que podem ser ordenadas.

**Suposições.** Utilize amostras aleatórias independentes. O teste  $U$  de Mann-Whitney testa a igualdade de duas distribuições. Para usá-lo para testar as diferenças de localização entre duas distribuições, deve-se assumir que as distribuições têm a mesma forma.

Para Obter Testes de Duas Amostras Independentes

1. A partir dos menus, escolha:

**Analisar > Testes Não Paramétricos > Diálogos Anteriores > 2 Amostras Independentes ...**

2. Selecione uma ou mais variáveis numéricas.
3. Selecione uma variável de agrupamento e clique em **Definir Grupos** para dividir o arquivo em dois grupos ou amostras.

### ***Tipos de teste Duas amostras independentes***

**Tipo de teste.** Quatro testes estão disponíveis para testar se duas amostras independentes (grupos) vêm da mesma população.

O **teste U de Mann-Whitney** é o mais popular dos testes de duas amostras independentes. Ele é equivalente ao teste de soma de ranqueamento Wilcoxon e ao teste de Kruskal-Wallis para dois grupos. O Mann-Whitney testa que duas populações amostradas são equivalentes em localização. As observações de ambos os grupos são combinadas e classificadas, com o ranqueamento médio designado no caso de empates. O número de empates deve ser pequeno em relação ao número total de observações. Se as populações são idênticas em localização, as classificações devem ser aleatoriamente combinadas entre as duas amostras. O teste calcula o número de vezes que uma pontuação do grupo 1 precede uma pontuação do grupo 2 e o número de vezes que uma pontuação do grupo 2 precede uma pontuação do grupo 1. A estatística *U* de Mann-Whitney é o menor desses dois números. A estatística *W* da soma de ranqueamento do Wilcoxon também é exibida. *W* é a soma das classificações do grupo com a classificação média menor, a menos que os grupos tenham a mesma classificação média; nesse caso, é a soma da classificação do grupo que é nomeado por último na caixa de diálogo Definir Grupos de Duas Amostras Independentes.

O **teste Z de Kolmogorov-Smirnov** e o **teste de sequências de Wald-Wolfowitz** são testes mais gerais que detectam diferenças em ambas as localizações e formas das distribuições. O teste de Kolmogorov-Smirnov baseia-se na diferença absoluta máxima entre as funções de distribuição acumulativas observadas para ambas as amostras. Quando essa diferença é significativamente grande, as duas distribuições são consideradas diferentes. O teste de sequências de Wald-Wolfowitz combina e classifica as observações de ambos os grupos. Se as duas amostras forem da mesma população, os dois grupos devem ser espalhados aleatoriamente ao longo do ranking.

O **teste de reações extremas de Moses** assume que a variável experimental afetará alguns assuntos em uma direção e outros assuntos na direção oposta. Os testes testam para obter respostas extremas em comparação a um grupo de controle. Este teste se concentra no intervalo do grupo de controle e é uma medida de quanto valores extremos no grupo experimental influenciam o intervalo quando combinados com o grupo de controle. O grupo de controle é definido pelo valor do grupo 1 na caixa de diálogo Grupos de definição de duas amostras independentes. Observações de ambos os grupos são combinadas e classificadas. O intervalo do grupo de controle é calculado como a diferença entre as classificações dos maiores e menores valores no grupo de controle mais 1. Como os outliers aleatórios podem distorcer facilmente o alcance do span, 5% dos casos de controle são cortados automaticamente de cada extremidade.

### ***Definir Grupos para Testes de Duas Amostras Independentes***

Para dividir o arquivo em dois grupos ou amostras, insira um valor inteiro para o Grupo 1 e outro valor para o Grupo 2. Casos com outros valores são excluídos da análise.

### ***Opções de Testes de Duas Amostras Independentes***

**Estatísticas.** É possível escolher uma ou ambas as estatísticas básicas.

- **Descritivo.** Exibe a média, o desvio padrão, o mínimo, o máximo e o número de casos não desconhecidos.
- **Quartis.** Exibe os valores correspondentes aos percentis 25, 50 e 75.

**Valores omissos.** Controla o tratamento de valores ausentes.

- **Excluir casos de teste por teste.** Quando vários testes são especificados, cada teste é avaliado separadamente para valores ausentes.
- **Excluir listwise dos casos.** Casos com valores ausentes para qualquer variável são excluídos de todas as análises.

## **Recursos adicionais do comando NPAR TESTS (Testes de duas amostras independentes)**

A linguagem de sintaxe de comando também permite especificar o número de casos a serem aparados para o teste de Moses (com o subcomando MOSES).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## **Testes de duas amostras relacionadas**

O procedimento Testes de duas amostras relacionadas compara as distribuições de duas variáveis.

**Exemplo.** Em geral, as famílias recebem o preço pedido quando vendem suas casas? Ao aplicar o teste de classificação assinada de Wilcoxon aos dados de 10 casas, você pode descobrir que sete famílias recebem menos do que o preço pedido, uma família recebe mais do que o preço pedido e duas famílias recebem o preço pedido.

**Estatísticas.** Média, desvio padrão, mínimo, máximo, número de casos não omissos e quartis. Testes: classificação assinada de Wilcoxon, sinal, McNemar. Se a opção Testes Exatos estiver instalada (disponível apenas em sistemas operacionais Windows), o teste de homogeneidade marginal também estará disponível.

Considerações de Dados de Testes de Duas Amostras Relacionadas

**Dados.** Use variáveis numéricas que podem ser ordenadas.

**Suposições.** Embora nenhuma distribuição particular seja assumida para as duas variáveis, a distribuição populacional das diferenças emparelhadas é assumida como simétrica.

Para Obter Testes de Duas Amostras Relacionadas

1. A partir dos menus, escolha:

**Analisar > Testes Não Paramétricos > Diálogos Anteriores > 2 Amostras Relacionadas ...**

2. Selecione um ou mais pares de variáveis.

## **Típos de teste de duas amostras relacionadas**

Os testes nesta seção comparam as distribuições de duas variáveis relacionadas. O teste apropriado a ser usado depende do tipo de dados.

Se seus dados forem contínuos, use o teste de sinal ou o teste de classificação sinalizada de Wilcoxon. O **teste de sinal** calcula as diferenças entre as duas variáveis para todos os casos e classifica as diferenças como positivas, negativas ou empatadas. Se as duas variáveis forem distribuídas de forma semelhante, o número de diferenças positivas e negativas não diferirá significativamente. O **teste dos postos sinalizados de Wilcoxon** considera informações sobre o sinal das diferenças e a magnitude das diferenças entre os pares. Como o teste de classificação sinalizada de Wilcoxon incorpora mais informações sobre os dados, ele é mais poderoso do que o teste de sinal.

Se seus dados são binários, use o **teste de McNemar**. Esse teste é normalmente usado em uma situação de medidas repetidas, na qual a resposta de cada sujeito é eliciada duas vezes, uma vez antes e uma vez após a ocorrência de um determinado evento. O teste de McNemar determina se a taxa de resposta inicial (antes do evento) é igual à taxa de resposta final (após o evento). Este teste é útil para detectar mudanças nas respostas devido à intervenção experimental em desenhos antes e depois.

Se seus dados são categóricos, use o **teste de homogeneidade marginal**. Este teste é uma extensão do teste de McNemar de resposta binária para resposta multinomial. Ele testa as mudanças na resposta (usando a distribuição qui-quadrado) e é útil para detectar mudanças na resposta devido à intervenção experimental em projetos de antes e depois. O teste de homogeneidade marginal está disponível apenas se você tiver instalado os Testes Exatos.

## **Opções de Testes de Duas Amostras Relacionadas**

**Estatísticas.** É possível escolher uma ou ambas as estatísticas básicas.

- **Descritivo.** Exibe a média, desvio padrão, mínimo, máximo e o número de casos não ausentes.
- **Quartis.** Exibe os valores correspondentes aos percentis 25, 50 e 75.

**Valores omissos.** Controla o tratamento de valores ausentes.

- **Excluir casos de teste por teste.** Quando vários testes são especificados, cada teste é avaliado separadamente para valores ausentes.
- **Excluir listwise dos casos.** Casos com valores ausentes para qualquer variável são excluídos de todas as análises.

### **Recursos Adicionais do Comando NPAR TESTS (Teste Qui-Quadrado)**

A linguagem de sintaxe de comando também permite testar uma variável com cada variável em uma lista.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## **Testes para diversas amostras independentes**

O procedimento Testes para Várias Amostras Independentes compara dois ou mais grupos de casos em uma variável.

**Exemplo.** Três marcas de lâmpadas de 100 watts diferem no tempo médio de vida em que elas irão queimar? Na análise de variância para um fator de Kruskal-Wallis, é possível ver que as três marcas diferem no tempo médio de vida.

**Estatísticas.** Média, desvio padrão, mínimo, máximo, número de casos não desconhecidos e quartis. Testes:  $H$  de Kruskal-Wallis, mediana.

Considerações de Dados de Testes para Várias Amostras Independentes

**Dados.** Utilize variáveis numéricas que possam ser ordenadas.

**Suposições.** Utilize amostras aleatórias independentes. O teste  $H$  de Kruskal-Wallis requer que as amostras testadas sejam semelhantes na forma.

Para Obter Testes para Várias Amostras Independentes

1. A partir dos menus, escolha:

**Analisar > Testes Não Paramétricos > Diálogos Anteriores > K Amostras Independentes ...**

2. Selecione uma ou mais variáveis numéricas.
3. Selecione uma variável de agrupamento e clique em **Definir Intervalo** para especificar valores de número inteiro mínimo e máximo para a variável de agrupamento.

### **Testes para vários tipos de teste de amostras independentes**

Três testes estão disponíveis para determinar se várias amostras independentes vêm da mesma população. O teste  $H$  de Kruskal-Wallis, o teste de mediana e o Jonckheere-Terpstra, testam se as várias amostras independentes são da mesma população.

O **teste H de Kruskal-Wallis H**, uma extensão do teste  $U$  de Mann-Whitney, é o análogo não paramétrico da análise de variância para um fator que detecta as diferenças na localização da distribuição. O **teste de mediana**, que é um teste mais geral (mas não tão poderosa), detecta as diferenças de distribuição no local e formato. O teste  $H$  de Kruskal-Wallis e o teste de mediana assumem que não ha ordenação *a priori* das populações  $k$  das quais as amostras são colhidas.

Quando *houver* uma ordenação natural *a priori* (crescente ou decrescente) das populações  $k$ , o **teste Jonckheere-Terpstra** será mais poderoso. Por exemplo, as populações  $k$  podem representar  $k$  aumentando as temperaturas. A hipótese de que temperaturas diferentes produzem a mesma distribuição de resposta é testada contra a alternativa de que à medida que a temperatura aumenta, a magnitude da resposta aumenta. Aqui, a hipótese alternativa é ordenada; portanto, Jonckheere-Terpstra é o teste mais apropriado para uso. O teste Jonckheere-Terpstra está disponível apenas se você instalou o módulo complementar Testes exatos.



## **Definir intervalo para testes de diversas amostras independentes**

Para definir o intervalo, insira valores de número inteiro para **Mínimo** e **Máximo** que correspondam às categorias mais baixa e mais alta da variável de agrupamento. Casos com valores fora dos limites são excluídos. Por exemplo, se especificar um valor mínimo de 1 e um valor máximo de 3, apenas os valores de número inteiro de 1 a 3 serão utilizados. O valor mínimo deve ser menor que o valor máximo, e ambos os valores devem ser especificados.

## **Opções de Testes para Várias Amostras Independentes**

**Estatísticas.** É possível escolher uma ou ambas as estatísticas básicas.

- **Descritivo.** Exibe a média, desvio padrão, mínimo, máximo e o número de casos não ausentes.
- **Quartis.** Exibe os valores correspondentes aos percentis 25, 50 e 75.

**Valores omissos.** Controla o tratamento de valores ausentes.

- **Excluir casos de teste por teste.** Quando vários testes são especificados, cada teste é avaliado separadamente para valores ausentes.
- **Excluir listwise dos casos.** Casos com valores ausentes para qualquer variável são excluídos de todas as análises.

## **Recursos Adicionais do Comando NPAR TESTS (K Amostras Independentes)**

A linguagem de sintaxe de comando também permite especificar um valor diferente da mediana observada para o teste de mediana (com o subcomando MEDIAN).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## **Testes para várias amostras relacionadas**

O procedimento Testes para Várias Amostras Relacionadas compara as distribuições de duas ou mais variáveis.

**Exemplo.** O público associa diferentes quantias de prestígio a um médico, advogado, policial e professor? Dez pessoas são solicitadas a classificar estas quatro ocupações em ordem de prestígio. O teste de Friedman indica que o público associa diferentes quantias de prestígio a essas quatro profissões.

**Estatísticas.** Média, desvio padrão, mínimo, máximo, número de casos não desconhecidos e quartis. Testes: Friedman, *W* de Kendall e *Q* de Cochran.

Considerações de Dados de Testes para Várias Amostras Relacionadas

**Dados.** Utilize variáveis numéricas que possam ser ordenadas.

**Suposições.** Testes não paramétricos não requerem suposições sobre a forma da distribuição subjacente. Utilize amostras aleatórias dependentes.

Para Obter Testes para Várias Amostras Dependentes

1. A partir dos menus, escolha:

**Analisar > Testes Não Paramétricos > Diálogos Anteriores > K Amostras Relacionadas ...**

2. Selecione duas ou mais variáveis de teste numéricas.

## **Testes para diversos tipos de teste de amostras relacionadas**

Três testes estão disponíveis para comparar as distribuições de várias variáveis relacionadas.

O teste de **Friedman** é o equivalente não paramétrico de um design de medidas repetidas de uma amostra ou uma análise de variância de dois fatores com uma observação por célula. Friedman testa a hipótese nula de que *k* variáveis relacionadas vêm da mesma população. Para cada caso, as variáveis *k* são ranqueadas de 1 a *k*. A estatística de teste é baseada nessas classificações.

**W de Kendall** é uma normalização da estatística de Friedman. *W* de Kendall é interpretável como o coeficiente de concordância, que é uma medida de acordo entre os avaliadores. Cada caso é um juiz ou avaliador, e cada variável é um item ou pessoa que está sendo avaliada. Para cada variável, a soma dos ranqueamentos é calculada. *W* de Kendall varia entre 0 (nenhuma concordância) e 1 (concordância total).

**Q de Cochran** é idêntico ao teste de Friedman, mas é aplicável quando todas as respostas são binárias. Este teste é uma extensão do teste de McNemar para a situação de amostra-*k*. *Q* de Cochran testa a hipótese de que diversas variáveis dicotômicas relacionadas têm a mesma média. As variáveis são medidas no mesmo indivíduo ou em indivíduos correspondidos.

### **Testes para várias estatísticas de amostras relacionadas**

É possível escolher estatísticas.

- **Descritivo.** Exibe a média, desvio padrão, mínimo, máximo e o número de casos não ausentes.
- **Quartis.** Exibe os valores correspondentes aos percentis 25, 50 e 75.

### **Recursos Adicionais do Comando NPAR TESTS (K Amostras Relacionadas)**

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## **Análise de Múltiplas Respostas**

---

### **Análise de Múltiplas Respostas**

Dois procedimentos estão disponíveis para analisar conjuntos de dicotomias múltiplas e de categorias múltiplas. O procedimento Frequências de Múltiplas Respostas exibe tabelas de frequência. O procedimento Crosstab de Múltiplas Respostas exibe tabulações cruzadas bidimensionais e tridimensionais. Antes de utilizar qualquer dos procedimentos, deve-se definir conjuntos de múltiplas respostas.

**Exemplo.** Este exemplo ilustra o uso de itens de múltiplas respostas em uma enquete de pesquisa de mercado. Os dados são fictícios e não devem ser interpretados como reais. Uma companhia aérea pode realizar uma pesquisa de opinião com passageiros que viajam em uma rota específica para avaliar as empresas concorrentes. Neste exemplo, a American Airlines deseja saber se seus passageiros também utilizam outras companhias aéreas para a rota Chicago - Nova York e se os horários e serviços são relativamente importantes na escolha de uma companhia aérea. Um comissário de bordo entrega para cada passageiro um breve questionário no momento do embarque. A primeira pergunta é: Circule todas as companhias aéreas com as quais você viajou pelo menos uma vez nos últimos seis meses nesta rota - American, United, TWA, USAir, Outra. Esta é uma pergunta de múltiplas respostas, já que o passageiro poderá circular mais de uma resposta. No entanto, essa pergunta não pode ser codificada diretamente porque uma variável pode ter somente um valor para cada caso. Deve-se utilizar diversas variáveis para mapear as respostas para cada pergunta. Existem duas maneiras de se fazer isso. Uma é definir uma variável que corresponda a cada uma das opções (por exemplo, American, United, TWA, USAir e Outra). Se o passageiro circula a United, a variável *united* é designada um código de 1, caso contrário 0. Este é um **método de dicotomia múltipla** de variáveis de mapeamento. A outra maneira de mapear as respostas é o **método de múltiplas categorias**, em que você estima o número máximo de respostas possíveis para a pergunta e configura o mesmo número de variáveis, com códigos utilizados para especificar a companhia aérea utilizada. Ao analisar uma amostra de questionários, é possível descobrir que nenhum usuário utilizou mais de três companhias aéreas diferentes nesta rota nos últimos seis meses. Além disso, você descobre que, devido à desregulamentação de companhias aéreas, 10 outras empresas foram incluídas na categoria Outro. Utilizando o método de múltiplas respostas, você define três variáveis, cada uma codificada como 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta*, e assim por diante. Se um determinado passageiro circular American e TWA, a primeira variável terá um código de 1, a segunda terá um código de 3, e a terceira terá um código de valor omisso. Outro passageiro pode ter circulado American e inserido Delta. Assim, a primeira variável terá um código de 1, a segunda possuirá um código de 5, e a terceira um código de valor omisso. Se você utilizar o método de múltiplas dicotomias, por outro lado, acabará tendo 14 variáveis separadas. Embora qualquer método de mapeamento seja viável para essa pesquisa de opinião, o método que você escolher dependerá da distribuição das respostas.

## Definir Conjuntos de Múltiplas Respostas

O procedimento Definir conjuntos de múltiplas respostas agrupa variáveis elementares em múltiplas dicotomias e conjuntos de múltiplas categorias, para os quais é possível obter tabelas de frequência e tabelas cruzadas. É possível definir até 20 conjuntos de múltiplas respostas. Cada conjunto deve ter um nome exclusivo. Para remover um conjunto, destaque-o na lista de conjuntos de múltiplas respostas e clique em **Remover**. Para alterar um conjunto, destaque-o na lista, modifique quaisquer características de definições de conjunto e clique em **Alterar**.

É possível codificar suas variáveis elementares como dicotomias ou categorias. Para utilizar variáveis dicotômicas, selecione **Dicotomias** para criar um conjunto de múltiplas dicotomias. Insira um valor de número inteiro para o valor Contado. Cada variável que possuir pelo menos uma ocorrência do valor contado se torna uma categoria do conjunto de múltiplas dicotomias. Selecione **Categorias** para criar um conjunto de categorias múltiplas com o mesmo intervalo de valores como variáveis de componente. Insira valores inteiros para os valores mínimo e máximo do intervalo para categorias do conjunto de várias categorias. O procedimento totaliza cada valor de número inteiro distinto no intervalo inclusivo em todas as variáveis do componente. As categorias vazias não são tabuladas.

Cada conjunto de múltiplas respostas deve receber um nome exclusivo de até sete caracteres. O procedimento prefixa um cifrão (\$) ao nome que você atribui. Não é possível usar os seguintes nomes reservados: *casenum*, *sysmis*, *jdate*, *date*, *time*, *length* e *width*. O nome do conjunto de múltiplas respostas existe apenas para uso em procedimentos de múltiplas respostas. Não é possível fazer referência a nomes de conjunto de múltiplas respostas em outros procedimentos. Opcionalmente, é possível inserir um rótulo de variável descritiva para o conjunto de múltiplas respostas. O rótulo pode ter até 40 caracteres de comprimento.

Para definir conjuntos de múltiplas respostas

1. A partir dos menus, escolha:

**Analisar > Múltiplas Respostas > Defina Conjuntos de Variáveis...**

2. Selecione duas ou mais variáveis.

3. Se suas variáveis forem codificadas como dicotomias, indique qual valor você deseja que seja contado. Se suas variáveis forem codificadas como categorias, defina o intervalo das categorias.

4. Insira um nome exclusivo para cada conjunto de múltiplas respostas.

5. Clique em **Incluir** para incluir o conjunto de múltiplas respostas na lista de conjuntos definidos.

## Frequências de múltiplas respostas

O procedimento de frequências de múltiplas respostas produz tabelas de frequência para conjuntos de múltiplas respostas. Deve-se primeiro definir um ou mais conjuntos de múltiplas respostas (consulte "Conjuntos de definição de múltiplas respostas").

Para múltiplos conjuntos de dicotomias, os nomes das categorias mostrados na saída vêm de rótulos de variáveis definidos para variáveis elementares no grupo. Se os rótulos das variáveis não estiverem definidos, os nomes das variáveis serão usados como rótulos. Para vários conjuntos de categorias, os rótulos de categoria vêm dos rótulos de valor da primeira variável do grupo. Se as categorias ausentes para a primeira variável estiverem presentes para outras variáveis no grupo, defina um rótulo de valor para as categorias ausentes.

**Valores omissos.** Os casos com valores ausentes são excluídos tabela por tabela. Alternativamente, é possível escolher um ou ambos os seguintes:

- **Excluir listwise dos casos dentro das dicotomias.** Exclui casos com valores ausentes para qualquer variável da tabulação do conjunto de dicotomias múltiplas. Isso se aplica apenas a conjuntos de múltiplas respostas definidos como conjuntos de dicotomia. Por padrão, um caso é considerado ausente para um conjunto de dicotomias múltiplas se nenhuma de suas variáveis de componente contiver o valor contado. Casos com valores ausentes para algumas (mas não todas as variáveis) são incluídos nas tabulações do grupo se pelo menos uma variável contiver o valor contado.

- **Excluir listwise dos casos dentro das categorias.** Exclui casos com valores ausentes para qualquer variável da tabulação do conjunto de múltiplas categorias. Isso se aplica apenas a conjuntos de múltiplas respostas definidos como conjuntos de categorias. Por padrão, um caso é considerado ausente para um conjunto de várias categorias somente se nenhum de seus componentes tiver valores válidos dentro do intervalo definido.

**Exemplo.** Cada variável criada a partir de uma pergunta de pesquisa é uma variável elementar. Para analisar um item de múltiplas respostas, você deve combinar as variáveis em um dos dois tipos de conjuntos de múltiplas respostas: um conjunto de múltiplas dicotomias ou um conjunto de múltiplas categorias. Por exemplo, se uma pesquisa de opinião de uma companhia aérea perguntar com quais das três companhias (American, United e TWA) você viajou nos últimos seis meses e você utilizou as variáveis dicotômicas e definiu um **conjunto de múltiplas dicotomias**, cada uma das três variáveis no conjunto se tornará uma categoria da variável de grupo. As contagens e porcentagens para as três companhias aéreas são exibidas em uma tabela de frequência. Se você descobrir que nenhum respondente mencionou mais de duas companhias aéreas, você pode criar duas variáveis, cada uma com três códigos, um para cada companhia aérea. Se você definir um **conjunto de categorias múltiplas**, os valores serão tabulados ao incluir os mesmos códigos nas variáveis elementares juntas. O conjunto de valores resultante é o mesmo para cada uma das variáveis elementares. Por exemplo, 30 respostas para o United são a soma das cinco respostas Unidas para a companhia aérea 1 e as 25 respostas Unidas para a companhia aérea 2. As contagens e porcentagens para as três companhias aéreas são exibidas em uma tabela de frequência.

**Estatísticas.** Tabelas de frequência exibindo contagens, porcentagens de respostas, porcentagens de casos, número de casos válidos e número de casos ausentes.

Considerações sobre dados de frequências de múltiplas respostas

**Dados.** Use conjuntos de múltiplas respostas.

**Suposições.** As contagens e porcentagens fornecem uma descrição útil para dados de qualquer distribuição.

**Procedimentos relacionados.** O procedimento Conjuntos de definição de múltiplas respostas permite definir conjuntos de múltiplas respostas.

Para Obter Frequências de Múltiplas Respostas

1. A partir dos menus, escolha:

**Analisar > Múltiplas Respostas > Frequências...**

2. Selecione um ou mais conjuntos de múltiplas respostas.

## Tabulações cruzadas de múltiplas respostas

O procedimento de tabulações cruzadas de múltiplas respostas efetua tabulações cruzadas de conjuntos de múltiplas respostas definidos, de variáveis elementares, ou uma combinação deles. Também é possível obter porcentagens de célula com base em casos ou respostas, modificar a manipulação de valores omissos ou obter tabulações cruzadas pairwise. Primeiro deve-se definir um ou mais conjuntos de múltiplas respostas (consulte "Para Definir Conjuntos de Múltiplas Respostas").

Para conjuntos de múltiplas dicotomias, os nomes de categoria mostrados na saída são provenientes de rótulos de variáveis definidos para variáveis elementares no grupo. Se os rótulos de variáveis não forem definidos, nomes de variáveis serão utilizados como rótulos. Para conjuntos de categorias múltiplas, os rótulos de categoria são provenientes de rótulos de valor da primeira variável no grupo. Se categorias omissas para a primeira variável estiverem presentes para outras variáveis no grupo, defina um rótulo de valor para as categorias omissas. O procedimento exibe rótulos de categoria para colunas em três linhas, com até oito caracteres por linha. Para evitar divisão de palavras, é possível reverter os itens de linha e de coluna ou redefinir rótulos.

**Exemplo.** Os conjuntos de múltiplas dicotomias e de múltiplas categorias podem efetuar tabulação cruzada com outras variáveis neste procedimento. Uma pesquisa de opinião para passageiros de companhias aéreas pede o seguinte: Circule todas as companhias aéreas com as quais você viajou pelo menos uma vez nos últimos seis meses (American, United, TWA). O que é mais importante na escolha de um voo - horário ou serviço? Selecione apenas um. Após inserir os dados como múltiplas dicotomias

ou categorias e combiná-las em um conjunto, é possível efetuar uma tabulação cruzada das opções da companhia aérea com a pergunta que envolve serviço ou horário.

**Estatísticas.** Tabulação cruzada com célula, linha, coluna e contagens totais, e com célula, linha, coluna e porcentagens totais. As porcentagens de célula podem ser baseadas em casos ou respostas.

Considerações de Dados para tabulações cruzadas de Múltiplas Respostas

**Dados.** Utilize conjuntos de múltiplas respostas ou variáveis categóricas numéricas.

**Suposições.** As contagens e porcentagens fornecem uma descrição útil dos dados a partir de qualquer distribuição.

**Procedimentos relacionados.** O procedimento Definir Conjuntos de Múltiplas Respostas permite definir conjuntos de múltiplas respostas.

Para Obter Crosstabs de Múltiplas Respostas

1. A partir dos menus, escolha:

**Analisar > Múltiplas Respostas > Crosstabs...**

2. Selecione uma ou mais variáveis numéricas ou conjuntos de múltiplas respostas para cada dimensão da tabulação cruzada.

3. Defina o intervalo de cada variável elementar.

Opcionalmente, é possível obter uma tabulação cruzada bidirecional para cada categoria de uma variável de controle ou conjunto de múltiplas respostas. Selecione um ou mais itens para a lista de Camada(s).

## Intervalos de Definição de Crosstabs de Múltiplas Respostas

Os intervalos de valores devem ser definidos para qualquer variável elementar na tabulação cruzada. Insira os valores de categoria mínimos e máximos inteiros que você deseja tabular. As categorias fora do intervalo são excluídas da análise. Os valores dentro do intervalo inclusivo são considerados inteiros (os não inteiros são truncados).

## Opções de Crosstabs de Múltiplas Respostas

**Porcentagens de Célula.** As contagens de célula são sempre exibidas. É possível optar por exibir porcentagens de linha, porcentagens de coluna e porcentagens de tabela de dois fatores (total).

**Porcentagens Baseadas em.** É possível basear porcentagens de célula em casos (ou respondentes). Isso não estará disponível se selecionar correspondência de variáveis em conjuntos de múltiplas categorias. Também é possível basear as porcentagens de célula em respostas. Para conjuntos de múltiplas dicotomias, o número de respostas é igual ao número de valores contabilizados nos casos. Para conjuntos de categorias múltiplas, o número de respostas é o número de valores no intervalo definido.

**Valores omissos.** É possível escolher uma ou ambas as seguintes opções:

- **Excluir listwise dos casos dentro das dicotomias.** Exclui casos com valores omissos para qualquer variável da tabulação do conjunto de múltiplas dicotomias. Isso se aplica apenas a conjuntos de múltiplas respostas definidos como conjuntos de dicotomias. Por padrão, um caso será considerado omissos para um conjunto de múltiplas dicotomias se nenhuma de suas variáveis de componente contiver o valor contado. Casos com valores omissos para algumas, mas não todas as variáveis são incluídas nas tabulações do grupo se pelo menos uma variável contiver o valor contado.
- **Excluir listwise dos casos dentro das categorias.** Exclui casos com valores omissos para qualquer variável da tabulação do conjunto de categorias múltiplas. Isso se aplica apenas a conjuntos de múltiplas respostas definidos como conjuntos de categoria. Por padrão, um caso será considerado omissos para um conjunto de categorias múltiplas apenas se nenhum de seus componentes possuir valores válidos dentro do intervalo definido.

Por padrão, ao efetuar tabulação cruzada de dois conjuntos de categorias múltiplas, o procedimento irá tabular cada variável no primeiro grupo com cada variável no segundo grupo e somará as contagens de

cada célula; portanto, algumas respostas podem aparecer mais de uma vez em uma tabela. É possível escolher a opção a seguir:

**Corresponder variáveis em conjuntos de respostas.** Emparelha a primeira variável do primeiro grupo com a primeira variável do segundo grupo, e assim por diante. Se essa opção for selecionada, o procedimento baseará as porcentagens da célula nas respostas e não nos respondentes. O emparelhamento não está disponível para conjuntos de múltiplas dicotomias ou variáveis elementares.

## Recursos adicionais do comando MULT RESPONSE

O idioma da sintaxe de comando também permite:

- Obter tabelas de tabulação cruzada com até cinco dimensões (com o subcomando BY).
- Alterar as opções de formatação de saída, incluindo a supressão de rótulos de valor (com o subcomando FORMAT).

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Resultados do relatório

---

### Resultados do Relatório

Listagens de casos e estatísticas descritivas são ferramentas básicas para estudar e apresentar dados. É possível obter listagens de caso com o Editor de Dados ou com o procedimento Sumarizar, contagens de frequência e estatísticas descritivas com o procedimento Frequências e as estatísticas de subpopulação com o procedimento Médias. Cada um desses utiliza um formato projetado para tornar as informações claras. Se desejar exibir as informações em um formato diferente, as Sumarizações de Relatório em Linhas e Sumarizações de Relatório em Colunas fornecem controle sobre a apresentação de dados.

### Sumarizações de Relatórios em Linhas

Os resumos de relatório em linhas produz relatórios nos quais diferentes estatísticas de resumo são dispostas em linhas. Listagens de casos também estão disponíveis, com ou sem estatísticas básicas.

**Exemplo.** Uma empresa com uma rede de lojas varejistas mantém registros de informações de funcionários, incluindo salário, tempo de serviço e a loja e divisão em que cada funcionário trabalha. É possível gerar um relatório que forneça informações de funcionário individual (listando), discriminado por loja e divisão (variáveis de quebra), com estatísticas básicas (por exemplo, média salarial) para cada loja, divisão e a divisão de cada loja.

**Colunas de Dados.** Lista as variáveis de relatório para as quais você deseja listagens de caso ou estatísticas básicas e controla o formato de exibição de colunas de dados.

**Colunas de Quebra.** Lista as variáveis de quebra opcionais que dividem o relatório em grupos e controla as estatísticas básicas e os formatos de exibição das colunas de quebra. Para diversas variáveis de quebra, haverá um grupo separado para cada categoria de cada variável de quebra dentro de categorias da variável de quebra anterior na lista. As variáveis de quebra devem ser variáveis categóricas discretas que dividem casos em um número limitado de categorias significativas. Valores individuais de cada variável de quebra aparecem, ordenados em uma coluna separada à esquerda de todas as colunas de dados.

**Relatório.** Controla as características gerais do relatório, incluindo estatísticas básicas gerais, exibição de valores omissos, numeração de página e títulos.

**Exibir casos.** Exibe os valores reais (ou rótulos de valor) as variáveis de coluna de dados para cada caso. Isso produz um relatório de listagem, que pode ser muito mais longo que um relatório sumarização.

**Visualização.** Exibe apenas a primeira página do relatório. Esta opção é útil para visualizar o formato de seu relatório sem processar o relatório inteiro.

**Dados já estão ordenados.** Para relatórios com variáveis de quebra, o arquivo de dados deve ser ordenado pelos valores de variável de quebra antes de gerar o relatório. Se o seu arquivo de

dados já estiver ordenado pelos valores de variáveis de quebra, será possível economizar tempo de processamento ao selecionar essa opção. Essa opção é útil principalmente após a execução de um relatório de visualização.

## Para obter um relatório de resumo: resumos em linhas

1. A partir dos menus, escolha:

**Analisar > Relatórios > Resumos de relatórios em linhas...**

2. Selecione uma ou mais variáveis para Colunas de Dados. Uma coluna no relatório é gerada para cada variável selecionada.
3. Para relatórios classificados e exibidos por subgrupos, selecione uma ou mais variáveis para colunas de quebra.
4. Para relatórios com estatísticas básicas para subgrupos definidos pelas variáveis de quebra, selecione a variável de quebra na lista de Variáveis de Colunas de Quebra e clique em **Sumarização** no grupo de Colunas de Quebra para especificar uma ou mais medidas de sumarização.
5. Para relatórios com estatísticas básicas gerais, clique em **Sumarização** para especificar uma ou mais medidas de sumarização.

## Formato de Quebra de Coluna de Dados do Relatório

As caixas de diálogo Formatar controlam os títulos das colunas, a largura das colunas, o alinhamento do texto e a exibição de valores de dados ou rótulos de valores. Formato da Coluna de Dados controla o formato das colunas de dados no lado direito da página do relatório. Formato de quebra controla o formato das colunas de quebra no lado esquerdo.

**Título da Coluna.** Para a variável selecionada, controla o título da coluna. Os títulos longos são agrupados automaticamente na coluna. Use a tecla Enter para inserir manualmente quebras de linha onde você deseja que os títulos sejam quebrados.

**Posição do Valor dentro da Coluna.** Para a variável selecionada, controla o alinhamento de valores de dados ou rótulos de valor dentro da coluna. O alinhamento de valores ou rótulos não afeta o alinhamento dos títulos das colunas. É possível recuar o conteúdo da coluna por um número especificado de caracteres ou centralizar o conteúdo.

**Conteúdo da Coluna.** Para a variável selecionada, controla a exibição de valores de dados ou rótulos de valores definidos. Os valores de dados são sempre exibidos para quaisquer valores que não tenham rótulos de valor definidos. (Não disponível para colunas de dados em relatórios de resumo de coluna.)

## Linhas de Sumarização de Relatório para/Linhas de Sumarização Final

As duas caixas de diálogo Linhas de Sumarização controlam a exibição das estatísticas básicas para grupos de quebra e para o relatório inteiro. Linhas de Sumarização controla estatísticas de subgrupo para cada categoria definida por uma ou mais variáveis de quebra. Linhas de Sumarização Final controla as estatísticas globais, exibidas no término do relatório.

As estatísticas básicas disponíveis são soma, média, mínimo, máximo, número de casos, porcentagem de casos acima ou abaixo de um valor especificado, porcentagem de casos dentro de um intervalo especificado de valores, desvio padrão, curtose, variância e assimetria.

## Opções de quebra de relatório

As opções de quebra controlam o espaçamento e a paginação das informações da categoria de quebra.

**Controle da Página.** Controla o espaçamento e a paginação das categorias da variável de quebra selecionada. É possível especificar um número de linhas em branco entre as categorias de quebra ou iniciar cada categoria de quebra em uma nova página.

**Linhas em Branco antes de Sumarizações.** Controla o número de linhas em branco entre os rótulos de categoria de quebra ou dados e estatísticas de resumo. Isso é particularmente útil para relatórios

combinados que incluem listagens de casos individuais e estatísticas resumidas para categorias de intervalo; nesses relatórios, é possível inserir espaço entre as listagens de casos e as estatísticas de resumo.

## Opções de relatório

As opções de relatório controlam o tratamento e a exibição de valores ausentes e a numeração da página do relatório.

**Excluir casos com listwise de valores omissos.** Elimina (do relatório) qualquer caso com valores ausentes para qualquer uma das variáveis do relatório.

**Valores Omissos Aparecem como.** Permite especificar o símbolo que representa os valores ausentes no arquivo de dados. O símbolo pode ser apenas um caractere e é utilizado para representar valores *omissos do sistema* e *omissos do usuário*.

**Numerar Páginas a partir de.** Permite especificar um número de página para a primeira página do relatório.

## Layout de relatórios

O Layout do Relatório controla a largura e o comprimento de cada página do relatório, o posicionamento do relatório na página e a inserção de linhas e rótulos em branco.

**Formato da Página.** Controla as margens da página expressas em linhas (superior e inferior) e caracteres (esquerda e direita) e informa o alinhamento dentro das margens.

**Títulos e Rodapés da Página.** Controla o número de linhas que separam os títulos e rodapés das páginas do corpo do relatório.

**Colunas de Quebra.** Controla a exibição de colunas de quebra. Se várias variáveis de quebra forem especificadas, elas poderão estar em colunas separadas ou na primeira coluna. Colocar todas as variáveis de quebra na primeira coluna produz um relatório mais restrito.

**Títulos da Coluna.** Controla a exibição dos títulos das colunas, incluindo o sublinhado do título, o espaço entre os títulos e o corpo do relatório e o alinhamento vertical dos títulos das colunas.

**Rótulos de Linhas e de Quebra de Colunas de Dados.** Controla o posicionamento das informações da coluna de dados (valores de dados e/ou estatísticas de resumo) em relação aos rótulos de quebra no início de cada categoria de quebra. A primeira linha de informações da coluna de dados pode começar na mesma linha que o rótulo da categoria de quebra ou em um número especificado de linhas após o rótulo da categoria de quebra. (Não disponível para relatórios de resumo de coluna.)

## Títulos de Relatório

Títulos de relatórios controla o conteúdo e o posicionamento dos títulos e rodapés dos relatórios. É possível especificar até 10 linhas de títulos de página e até 10 linhas de rodapés de página, com componentes justificados à esquerda, centralizados e justificados à direita em cada linha.

Se você inserir variáveis em títulos ou rodapés, o rótulo do valor atual ou o valor da variável será exibido no título ou rodapé. Nos títulos, é exibido o rótulo do valor correspondente ao valor da variável no início da página. Nos rodapés, é exibido o rótulo do valor correspondente ao valor da variável no final da página. Se não houver rótulo de valor, o valor real será exibido.

**Variáveis Especiais.** As variáveis especiais *DATE* e *PAGE* permitem inserir a data ou o número da página atual em qualquer linha de um cabeçalho ou rodapé de relatório. Se o arquivo de dados contiver variáveis denominadas *DATE* ou *PAGE*, não será possível utilizar essas variáveis em títulos ou rodapés do relatório.

## Resumos de relatórios em colunas

Resumos de relatórios em colunas produz relatórios de resumo nos quais diferentes estatísticas básicas aparecem em colunas separadas.



**Exemplo.** Uma empresa com uma rede de lojas varejistas mantém registros de informações de funcionários, incluindo salário, tempo de serviço e a divisão em que cada funcionário trabalha. Você pode gerar um relatório que forneça estatísticas salariais básicas (por exemplo, média, mínimo e máximo) para cada divisão.

**Colunas de Dados.** Lista as variáveis de relatório para as quais você deseja estatísticas básicas e controla o formato de exibição e as estatísticas básicas exibidas para cada variável.

**Colunas de Quebra.** Lista as variáveis de quebra opcionais que dividem o relatório em grupos e controla os formatos de exibição das colunas de quebra. Para várias variáveis de quebra, haverá um grupo separado para cada categoria de cada variável de quebra dentro das categorias da variável de quebra anterior na lista. As variáveis de quebra devem ser variáveis categóricas discretas que dividem os casos em um número limitado de categorias significativas.

**Relatório.** Controla as características gerais do relatório, incluindo a exibição de valores ausentes, numeração de páginas e títulos.

**Visualização.** Exibe apenas a primeira página do relatório. Esta opção é útil para visualizar o formato do seu relatório sem processar o relatório inteiro.

**Dados já estão ordenados.** Para relatórios com variáveis de quebra, o arquivo de dados deve ser classificado por valores de variável de quebra antes de gerar o relatório. Se o seu arquivo de dados já estiver ordenado por valores das variáveis de quebra, é possível economizar tempo de processamento selecionando esta opção. Essa opção é particularmente útil após a execução de um relatório de visualização.

## Para obter um relatório de resumo: resumos em colunas

1. No menu, escolha:

**Analisar > Relatórios > Resumos de relatórios em colunas...**

2. Selecione uma ou mais variáveis para Colunas de Dados. Uma coluna no relatório é gerada para cada variável selecionada.
3. Para alterar a medida de sumarização de uma variável, selecione a variável na lista Variáveis de Coluna de Dados e clique em **Sumarização**.
4. Para obter mais de uma medida de resumo para uma variável, selecione a variável na lista de origem e mova-a para a lista Variáveis da Coluna de Dados várias vezes, uma para cada medida de resumo desejada.
5. Para exibir uma coluna contendo a soma, média, razão ou outra função de colunas existentes, clique em **Inserir Total**. Isso coloca uma variável denominada *total* na lista Colunas de Dados.
6. Para relatórios classificados e exibidos por subgrupos, selecione uma ou mais variáveis para Colunas de quebra.

## Função de resumo das colunas de dados

Linhas de Sumarização controla a estatística de sumarização exibida para variável da coluna de dados selecionada.

As estatísticas básicas disponíveis são soma, média, mínimo, máximo, número de casos, porcentagem de casos acima ou abaixo de um valor especificado, porcentagem de casos dentro de um intervalo especificado de valores, desvio padrão, variância, curtose e assimetria.

## Sumarização de Colunas de Dados para Total de Colunas

A Coluna de Resumo controla o total de estatísticas de resumo que resumem duas ou mais colunas de dados.

As estatísticas básicas totais disponíveis são soma de colunas, média de colunas, mínimo, máximo, diferença entre valores em duas colunas, quociente de valores em uma coluna dividido por valores em outra coluna, e o produto dos valores de colunas multiplicados juntos.

**Soma de colunas.** A coluna *total* é a soma das colunas na lista Colunas de Sumarização.

**Média de colunas.** A coluna *total* é a média das colunas na lista Colunas de Sumarização.

**Mínimo de colunas.** A coluna *total* é o mínimo das colunas na lista de Colunas de Sumarização.

**Máximo de colunas.** A coluna *total* é o máximo das colunas na lista Colunas de Sumarização.

**1ª coluna - 2ª coluna.** A coluna *total* é a diferença das colunas na lista de Colunas de Sumarização. A lista de Coluna de Resumo deve conter exatamente duas colunas.

**1ª coluna / 2ª coluna.** A coluna *total* é o quociente das colunas na lista Colunas de Sumarização. A lista de Coluna de Resumo deve conter exatamente duas colunas.

**% 1ª coluna / 2ª coluna.** A coluna *total* é a porcentagem da primeira coluna da segunda coluna na lista Coluna de Sumarização. A lista de Coluna de Resumo deve conter exatamente duas colunas.

**Produto de colunas.** A coluna *total* é o produto das colunas na lista de Coluna de Sumarização.

## Formato da Coluna do Relatório

As opções de formatação de dados e coluna de quebra para Sumarizações de Relatórios em Colunas são as mesmas que aquelas descritas para Sumarizações de Relatórios em Linhas.

## Resumos de relatórios nas opções de quebra de colunas

As opções de quebra controlam a exibição do subtotal, o espaçamento e a paginação das categorias de quebra.

**Subtotal.** Controla os subtotais de exibição para categorias de quebra.

**Controle da Página.** Controla o espaçamento e a paginação das categorias da variável de quebra selecionada. É possível especificar um número de linhas em branco entre as categorias de quebra ou iniciar cada categoria de quebra em uma nova página.

**Linhas em Branco antes do subtotal.** Controla o número de linhas em branco entre os dados da categoria de quebra e os subtotais.

## Resumos de relatórios nas opções de colunas

Opções controla a exibição de totais gerais, a exibição de valores omissos e a paginação em relatórios sumarização de coluna.

**Total Geral.** Exibe e rotula um total geral para cada coluna; exibido na parte inferior da coluna.

**Valores omissos.** É possível excluir valores omissos do relatório ou selecionar um único caractere para indicar valores omissos no relatório.

## Layout de Relatório para Sumarizações nas colunas

As opções de layout de relatório para Sumarizações de Relatórios em Colunas são as mesmas que aquelas descritas para Sumarizações de Relatórios em Linhas.

## Recursos adicionais do comando REPORT

O idioma da sintaxe de comando também permite:

- Exiba diferentes funções de resumo nas colunas de uma única linha de resumo.
- Insira linhas de resumo em colunas de dados para variáveis que não sejam a variável da coluna de dados ou para várias combinações (funções compostas) de funções de resumo.
- Use Mediana, Modo, Frequência e Porcentagem como funções de resumo.
- Controle com mais precisão o formato de exibição das estatísticas básicas.
- Insira linhas em branco em vários pontos nos relatórios.

- Inserir linhas em branco após cada  $n$ -ésimo caso nos relatórios de listagem.

Devido à complexidade da sintaxe REPORT, talvez você ache útil, ao construir um novo relatório com a sintaxe, aproximar o relatório gerado das caixas de diálogo, copiar e colar a sintaxe correspondente e refinar essa sintaxe para produzir o relatório exato desejado.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Análise de Confiabilidade

---

A análise de confiabilidade permite estudar as propriedades das escalas de medida e os itens que compõem as escalas. O procedimento de Análise de Confiabilidade calcula uma série de medidas de Confiabilidade de escala comumente usadas e também fornece informações sobre as relações entre itens individuais na escala. Coeficientes de correlação intraclasse podem ser usados para calcular estimativas de confiabilidade entre avaliadores.

A análise de confiabilidade também fornece estatísticas de Kappa de múltiplos avaliadores da Fleiss que avaliam a concordância entre avaliadores para determinar a confiabilidade entre os vários avaliadores. Um acordo mais alto fornece mais confiança nas classificações que refletem a verdadeira circunstância. As opções de Kappa de avaliador múltiplo de Fleiss estão disponíveis no diálogo “Análise de confiabilidade: estatísticas” na página 228.

### Exemplo

O meu questionário mede a satisfação do cliente de maneira útil? Usando a análise de confiabilidade, você pode determinar até que ponto os itens do questionário estão relacionados entre si, obter um índice geral da repetibilidade ou consistência interna da escala como um todo e identificar itens problemáticos que devem ser excluídos da escala.

### Estatísticas

Descritivas para cada variável e para as estatísticas básicas, de escala em itens, correlações e covariâncias entre itens, estimativas de confiabilidade, tabela ANOVA, coeficientes de correlação intraclasse,  $T^2$  de Hotelling, teste de aditividade de Tukey e Kappa de avaliador múltiplo de Fleiss.

### Modelos

Os modelos de confiabilidade a seguir estão disponíveis:

#### Alfa (Cronbach)

Esse modelo é uma medida de consistência interna com base na correlação média entre itens.

#### Omega (McDonald's)

Este modelo assume que o modelo é unidimensional incluindo um único fator sem dependência de item local na forma de covariâncias de erro. O modelo implica que a covariância dos dois itens diferentes é o produto de seus carregamentos.

#### Divisão em metade

Esse modelo divide a escala em duas partes e examina a correlação entre as partes.

#### Guttman

Este modelo calcula os limites inferiores de Guttman para confiabilidade real.

#### Paralelo

Esse modelo supõe que todos os itens possuem variâncias e variâncias de erro iguais entre as replicações.

#### Paralelo-série

Esse modelo faz suposições do modelo Paralelo e também supõe médias iguais entre os itens.

## Considerações sobre dados de Análise de confiabilidade

### Dados

Os dados podem ser dicotômicos, ordinais ou intervalo, porém devem ser codificados numericamente.

## Suposições

As observações devem ser independentes, e os erros deverão estar não correlacionados entre os itens. Cada par de itens deve ter uma distribuição normal bivariada. As escalas devem ser aditivas, de modo que cada item esteja linearmente relacionado ao escore total. As suposições a seguir se aplicam às estatísticas de Kappa de Multiple Rater de Fleiss:

- Pelo menos duas variáveis de item devem ser selecionadas para executar qualquer estatística de confiabilidade.
- Quando pelo menos duas variáveis de classificação são selecionadas, a sintaxe de Kappa de Multiple Rater de Fleiss é colada.
- Não há nenhuma conexão entre avaliadores.
- O número de avaliadores é uma constante.
- Cada assunto é classificado pelo mesmo grupo contendo somente um avaliador único.
- Nenhuma ponderação pode ser designada às diversas discordâncias.

## Procedimentos Relacionados

Se você deseja explorar a dimensionalidade de seus itens de escala (para ver se mais de um construto é necessário para explicar o padrão de pontuação dos itens), use a análise fatorial ou o ajuste de escala multidimensional. Para identificar grupos homogêneos de variáveis, use a análise de cluster hierárquica para as variáveis de cluster.

## Para obter uma Análise de confiabilidade

1. A partir dos menus, escolha:

**Analisar > Escala > Análise de confiabilidade...**

2. Selecione duas ou mais variáveis como componentes em potencial de uma escala aditiva.

3. Escolha um modelo da lista suspensa **Modelo**.

4. Opcionalmente, clique em **Estatísticas** para selecionar várias estatísticas que descrevem seus itens de escala ou concordância entre avaliadores.

## Análise de confiabilidade: estatísticas

É possível selecionar várias estatísticas que descrevem sua escala, itens e a concordância entre os avaliadores para determinar a confiabilidade entre os vários avaliadores. As estatísticas relatadas por padrão incluem o número de casos, o número de itens e as estimativas de confiabilidade conforme a seguir:

### Modelos alpha

Coeficiente alpha; para dados dicotômicos, isso é equivalente ao coeficiente de Kuder-Richardson 20 (KR20).

### Modelos Omega

Estimação de omega do McDonald's para avaliar a confiabilidade.

### Modelos de duas metades

Correlação entre formulários, confiabilidade de Guttman split-half, confiabilidade de Spearman-Brown (comprimento igual e desigual) e coeficiente alpha para cada metade.

### Modelos de Guttman

Coeficientes de confiabilidade lambda 1 a lambda 6.

### Modelos paralelos e paralelos-série

Teste de qualidade de ajuste do modelo; estimativas de variância de erro, variância comum e variância verdadeira; correlação entre itens comum estimada; confiabilidade estimada; e estimativa imparcial de confiabilidade.

### Descritivos para

Produz estatísticas descritivas para escalas ou itens entre os casos.

**Item**

Produz estatísticas descritivas para os itens entre os casos.

**Escala**

Produz estatísticas descritivas para escalas.

**Ajuste de escala se o item for excluído**

Exibe estatísticas resumidas comparando cada item com a escala composta pelos outros itens. As estatísticas incluem média e variância da escala se o item for excluído da escala, correlação entre o item e a escala composta por outros itens e o Alfa de Cronbach se o item for excluído da escala.

**Sumários**

Fornecer estatísticas descritivas das distribuições do item em todos os itens na escala.

**Médias**

Estatísticas de sumarização para médias de item São exibidas as médias de item maiores, menores e médias, o intervalo e a variância das médias de item e a razão das médias de item, das maiores às menores.

**Variâncias**

Estatísticas de sumarização para variâncias de itens As variâncias de item maiores, menores e médias, o intervalo e a variância de variâncias de item e a razão das variâncias de item maiores com as menores são exibidos.

**Correlações**

Estatísticas de resumo para correlações entre itens. As correlações entre itens maiores, menores e médias, o intervalo e a variância de correlações entre itens e a razão das correlações entre itens maiores com as menores são exibidos.

**Covariâncias**

Estatísticas básicas para covariâncias entre itens. As covariâncias entre itens maiores, menores e médias, o intervalo e a variância de covariâncias entre itens e a razão das covariâncias entre itens maiores com as menores são exibidos.

**Interitem**

Produz matrizes de correlações ou covariâncias entre itens.

**Tabela ANOVA**

Produz testes de médias iguais.

**Teste F**

Exibe uma tabela de análise de variância de medidas repetidas.

**Qui-quadrado de Friedman**

Exibe o coeficiente de concordância de qui-quadrado e Kendall de Friedman. Essa opção é apropriada para os dados que estiverem na forma de ranqueamentos. O teste qui-quadrado substitui o teste F usual na tabela ANOVA.

**Qui-quadrado de Cochran**

Exibe o Q de Cochran. Esta opção é apropriada para dados que são dicotômicos. A estatística Q substitui a estatística F usual na tabela ANOVA.

**Concordância entre avaliadores: Kappa de Fleiss**

Avalia a concordância entre avaliadores para determinar a confiabilidade entre os diversos avaliadores. Uma concordância mais alta fornece mais confiança nas classificações que refletem a verdadeira circunstância. A estatística kappa não ponderada generalizada mede a concordância entre qualquer número constante de avaliadores, assumindo:

- Pelo menos duas variáveis de item devem ser especificadas para executar qualquer estatística de confiabilidade.
- Pelo menos duas variáveis de classificações devem ser especificadas.
- As variáveis selecionadas como itens também podem ser selecionadas como classificações.
- Não há nenhuma conexão entre avaliadores.
- O número de avaliadores é uma constante.
- Cada assunto é classificado pelo mesmo grupo contendo somente um avaliador único.

- Nenhuma ponderação pode ser designada às diversas discordâncias.

#### **Exibir acordo em categorias individuais**

Especifica se a concordância em categorias individuais será ou não exibida na saída. Por padrão, a saída suprime a estimativa em qualquer categoria individual. Quando ativado, múltiplas tabelas são exibidas na saída.

#### **Ignorar maiúsculas e minúsculas da sequência de caracteres**

Controla se as variáveis de sequência fazem ou não distinção entre maiúsculas e minúsculas. Por padrão, os valores de classificação da sequência de caracteres diferenciam maiúsculas de minúsculas.

#### **Os rótulos de categoria da sequência de caracteres são exibidos em letras maiúsculas**

Controla se os rótulos de categoria nas tabelas de saída são exibidos em maiúsculas ou minúsculas. A configuração está ativada por padrão, que exibe os rótulos de categoria de sequência de caracteres em letras maiúsculas.

#### **Nível de significância assintótico (%)**

Especifica o nível de significância para os intervalos de confiança assintóticos. 95 é a configuração padrão.

### **Ausente**

#### **Excluir valores ausentes do usuário e do sistema**

Controla a exclusão de valores omissos do usuário e do sistema. Por padrão, os valores ausentes do usuário e do sistema são excluídos.

#### **Os valores ausentes do usuário são tratados como válidos**

Quando ativada, trata os valores omissos do usuário e do sistema como dados válidos. A configuração está desativada por padrão.

### **Quadrado T de Hotelling**

Produz um teste multivariado da hipótese nula de que todos os itens na escala têm a mesma média.

### **Teste de aditividade de Tukey**

Produz um teste da suposição de que não há nenhuma interação multiplicativa entre os itens.

### **Coefficiente de correlação intraclasses**

Produz medidas de consistência ou concordância de valores dentro dos casos.

#### **Modelo**

Selecione o modelo para calcular o coeficiente de correlação intraclasses. Os modelos disponíveis são Misto de duas vias, Aleatório de duas vias e Aleatório de uma via. Selecione **Dois fatores combinados** quando os efeitos de pessoas são aleatórios e os efeitos de item são fixos, selecione **Dois fatores aleatórios** quando os efeitos de pessoas e os efeitos de item são aleatórios, ou selecione **Um fator aleatório** quando os efeitos de pessoas são aleatórios.

#### **Tipo**

Selecione o tipo de índice. Os tipos disponíveis são Consistência e Acordo Absoluto.

#### **Intervalo de confiança (%)**

Especifique o nível do intervalo de confiança. O padrão é 95%.

#### **Valor de teste**

Especifique o valor hipotético do coeficiente para o teste de hipótese. Este valor é o valor com o qual o valor observado é comparado. O valor padrão é 0.

## **Recursos Adicionais do Comando RELIABILITY**

O idioma da sintaxe de comando também permite:

- Ler e analisar uma matriz de correlações.
- Gravar uma matriz de correlações para análise posterior.
- Especificar divisões diferentes de metades iguais para o método das duas metades.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Kappa ponderado

---

O kappa ponderado de Cohen é amplamente utilizado na classificação cruzada como uma medida de concordância entre os avaliadores observados. É um índice de concordância apropriado quando as classificações são escalas nominais sem estrutura de ordem. O desenvolvimento do kappa ponderado de Cohen foi motivado pelo fato de que algumas atribuições em uma tabela de contingência podem ser de maior gravidade do que outras. A estatística baseia-se nos pesos de células predefinidos que refletem concordância ou discordância.

O procedimento Kappa ponderado fornece opções para estimar o kappa ponderado de Cohen, uma importante generalização da estatística kappa que mede a concordância de dois sujeitos ordinais com categorias idênticas.

**Nota:** O procedimento de Kappa ponderado substitui a funcionalidade fornecida anteriormente pela extensão `STATS WEIGHTED KAPPA.spe`.

### Exemplo

Existem situações em que as diferenças entre os avaliadores não devem ser tratadas como igualmente importantes. Um exemplo seria no setor de saúde, em que várias pessoas coletam dados de pesquisa ou clínicos. Nesses casos, a confiabilidade dos dados pode ser questionada, dada a variabilidade entre aqueles que coletam os dados.

### Estatísticas

Kappa ponderado de Cohen, escala linear, escala quadrática, intervalo de confiança assintótico.

## Considerações sobre dados do Kappa ponderado

### Dados

Uma tabela bidirecional baseada em um conjunto de dados ativo é necessária para estimar a estatística do kappa ponderado de Cohen.

As variáveis de classificação devem ser do mesmo tipo (todas de sequência ou todas numéricas).

A estimativa do kappa ponderado de Cohen só faz sentido quando as categorias das duas variáveis de classificação, representadas pela linha e coluna na tabela, são ordenadas adequadamente (para um par de variáveis numéricas, a ordem numérica é aplicada; para um par de variáveis de sequência de caracteres, a ordem alfabética é aplicada).

### Suposições

Quando pares de variáveis mistas são selecionados, o kappa ponderado de Cohen não é estimado.

Presume-se que as variáveis de classificação compartilhem o mesmo conjunto de categorias.

## Para obter uma análise do Kappa ponderado

1. A partir dos menus, escolha:

**Analisar > Escala > Kappa ponderado...**

2. Selecione duas ou mais variáveis de sequência ou numéricas para especificar como **Avaliadores pairwise**.

**Nota:** Deve-se selecionar todas as variáveis de sequência ou todas as variáveis numéricas.

3. Opcionalmente, ative a configuração **Especificar avaliadores para linhas e colunas** para controlar a exibição de avaliadores pairwise ou avaliadores de linhas/colunas.

- Quando ativada, os avaliadores pairwise são suprimidos e os avaliadores de linhas/colunas são exibidos. A interface com o usuário é atualizada para fornecer campos **Avaliador(es) de linhas** e **Avaliador(es) de colunas** (substituindo efetivamente o campo **Avaliadores pairwise**).
- Quando desativado, os avaliadores de linha/coluna são suprimidos e os avaliadores de pares são exibidos (a configuração padrão)

Quando **Especificar avaliadores para linhas e colunas** estiver ativado, especifique pelo menos uma variável para **Avaliador(es) de linhas** e **Avaliador(s) de colunas**.

**Nota:** Se **Avaliador(es) de linhas** e **Avaliador(s) de colunas** contiverem somente uma variável, as variáveis selecionadas não poderão ser as mesmas para ambos.

4. Opcionalmente, clique em **Crítérios** para especificar as configurações de escala de ponderação e de valores omissos ou em **Imprimir** para especificar as configurações de formato de exibição e de tabulação cruzada.

## Kappa ponderado: Critérios

O diálogo Critérios fornece opções para especificar a estimação das estatísticas kappa ponderado de Cohen.

### Escala de ponderação

Fornece opções para especificar uma ponderação linear ou quadrática para concordância. O uso de peso linear (também conhecido como Cicchetti-Allison) é a configuração padrão.

### Valores omissos

Fornece opções para remover casos com valores omissos em uma base pairwise e tratar valores omissos do usuário como válidos.

### As variáveis de classificação de sequência fazem distinção entre maiúsculas e minúsculas

Quando selecionada, as variáveis de sequência são tratadas como tendo distinção entre maiúsculas e minúsculas.

### Intervalo de confiança assintótico (%)

Esta configuração opcional especifica o nível de confiança para a estimação dos intervalos de confiança assintóticos. Deve ser um valor duplo único entre 0 e 100 (95 é a configuração padrão).

## Kappa ponderado: imprimir

O diálogo Impressão fornece opções para controlar as tabelas de tabulação cruzada.

### Exibição e formato

Fornece opções para controlar a exibição e o formato da tabela de tabulação cruzada.

#### As categorias de classificação são exibidas em ordem crescente

Quando selecionadas, as categorias de classificação nas tabelas de tabulação cruzada são exibidas em ordem crescente. Quando a configuração não é selecionada, as categorias de classificação nas tabelas de tabulação cruzada são exibidas em ordem decrescente. Essa configuração é ativada por padrão.

#### Os rótulos de categoria de string são exibidos em letras maiúsculas

Quando selecionada, as tabelas de tabulação cruzada são exibidas em maiúsculas. Quando a configuração não está selecionada, as tabelas de tabulação cruzada são exibidas em letras minúsculas. Essa configuração é ativada por padrão.

### Tabulação cruzada

Fornece opções para índice as variáveis de classificação que são usadas em tabulação cruzada. Por padrão, as configurações de tabulação cruzada não estão habilitadas, o que suprime a tabulação cruzada de quaisquer variáveis de classificação.

#### Exibir a tabulação cruzada de variáveis de classificação

Quando selecionada, essa configuração permite a tabulação cruzada de todas as variáveis de classificação especificadas pelo usuário.

#### Incluir todas as variáveis de classificação

Quando selecionada, as tabelas de tabulação cruzada são impressas para todos os pares de variáveis de classificação definidos.

#### Incluir variáveis de classificação especificadas pelo usuário

Quando selecionada, use os campos **Variáveis disponíveis**, **Avaliadores de linhas** e **Avaliadores de colunas** para selecionar quais variáveis de classificação serão incluídas nas tabelas de tabulação cruzada.



## Ajuste de escala multidimensional

---

O dimensionamento multidimensional tenta encontrar a estrutura em um conjunto de medidas de distância entre objetos ou casos. Essa tarefa é realizada atribuindo observações a locais específicos em um espaço conceitual (geralmente bi ou tridimensional) de modo que as distâncias entre os pontos no espaço correspondam às diferenças dadas o mais próximo possível. Em muitos casos, as dimensões desse espaço conceitual podem ser interpretadas e usadas para entender melhor seus dados.

Se você tiver variáveis medidas objetivamente, poderá usar o dimensionamento multidimensional como uma técnica de redução de dados (o procedimento Multidimensional Scaling calculará as distâncias dos dados multivariados para você, se necessário). O ajuste de escala multidimensional também pode ser aplicado a classificações subjetivas de dissimilaridade entre objetos ou conceitos. Além disso, o procedimento Ajuste de escala multidimensional pode manipular dados de dissimilaridade de várias fontes, como você pode fazer com vários avaliadores ou respondentes do questionário.

**Exemplo.** Como as pessoas percebem as relações entre carros diferentes? Se você tiver dados de entrevistados indicando classificações de similaridade entre diferentes marcas e modelos de carros, o ajuste de escala multidimensional pode ser usada para identificar dimensões que descrevem as percepções dos consumidores. É possível descobrir, por exemplo, que o preço e o tamanho de um veículo definem um espaço bidimensional, o que explica as semelhanças relatadas por seus entrevistados.

**Estatísticas.** Para cada modelo: matriz de dados, matriz de dados idealmente escalada, estresse de S (Young), estresse (Kruskal), RSQ, coordenadas de estímulo, estresse médio e RSQ para cada estímulo (modelos RMDS). Para modelos de diferenças individuais (INDSCAL): pesos de sujeitos e índice de estranheza para cada sujeito. Para cada matriz em modelos de ajuste de escala multidimensional replicados: estresse e RSQ para cada estímulo. Gráficos: coordenadas de estímulo (bi ou tridimensionais), gráfico de dispersão de disparidades versus distâncias.

Considerações de dados para ajuste de escala multidimensional

**Dados.** Se seus dados são dados de dissimilaridade, todas as dissimilaridades devem ser quantitativas e devem ser medidas na mesma métrica. Se seus dados forem dados multivariados, as variáveis podem ser dados quantitativos, binários ou de contagem. O ajuste de escala de variáveis é uma questão importante - as diferenças no ajuste de escala podem afetar sua solução. Se suas variáveis tiverem grandes diferenças de escala (por exemplo, uma variável é medida em dólares e a outra variável é medida em anos), considere padronizá-las (esse processo pode ser feito automaticamente pelo procedimento Ajuste de escala multidimensional).

**Suposições.** O procedimento Ajuste de escala multidimensional é relativamente livre de suposições distributivas. Certifique-se de selecionar o nível de medição apropriado (ordinal, intervalo ou razão) na caixa de diálogo Opções de ajuste de escala multidimensional para que os resultados sejam calculados corretamente.

**Procedimentos relacionados.** Se seu objetivo for a redução de dados, um método alternativo a ser considerado é a análise fatorial, principalmente se suas variáveis forem quantitativas. Se desejar identificar grupos de casos semelhantes, considere complementar o ajuste de escala multidimensional com uma análise de cluster hierárquico ou por  $k$ -médias.

Para Obter uma Análise de Ajuste de Escala Multidimensional

1. A partir dos menus, escolha:

**Analisar > Escala > Ajuste de Escala Multidimensional...**

2. Selecione pelo menos quatro variáveis numéricas para análise.

3. No grupo Distâncias, selecione **Dados são distâncias** ou **Criar distâncias a partir de dados**.

4. Se selecionar **Criar distâncias a partir de dados**, também será possível selecionar uma variável de agrupamento para matrizes individuais. A variável de agrupamento pode ser numérica ou sequência de caracteres.

Opcionalmente, é possível:

- Especifique a forma da matriz de distância quando os dados são distâncias.

- Especifique a medida de distância a ser usada ao criar distâncias de dados.

## Forma de dados para ajuste de escala multidimensional

Se o seu conjunto de dados ativo representar distâncias entre um conjunto de objetos ou representar distâncias entre dois conjuntos de objetos, especifique a forma de sua matriz de dados para obter os resultados corretos.

*Nota:* Não será possível selecionar **Quadrado simétrico** se a caixa de diálogo Modelo especificar condicionalidade de linha.

## Criação de Medida para Ajuste de Escala Multidimensional

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 no grupo Medida correspondente ao seu tipo de dados e, em seguida, escolha uma das medidas na lista suspensa correspondente ao tipo de medida. As alternativas disponíveis são:

- **Intervalo.** Distância euclidiana, distância euclidiana quadrada, Chebychev, bloco, Minkowski ou customizado.
- **Contagens.** 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 certos casos, como quando as variáveis são medidas em escalas muito diferentes, convém padronizar os valores antes de calcular as proximidades (não aplicável a dados binários). Escolha um método de padronização a partir da lista suspensa Padronizar. Se nenhuma padronização for necessária, escolha **Nenhum**.

## Modelo de Ajuste de Escala Multidimensional

A estimação de correção de um modelo de ajuste de escala multidimensional depende dos aspectos dos dados e do próprio modelo.

**Nível de Medição.** Permite especificar o nível de seus dados. As alternativas são Ordinal, Intervalo ou Razão. Se as variáveis forem ordinais, selecionar **Desempatar observações empatadas** solicita que as variáveis sejam tratadas como variáveis contínuas, de modo que os empates (valores iguais para casos diferentes) sejam resolvidos corretamente.

**Condicionalidade.** Permite especificar quais comparações são significativas. As alternativas são Matriz, Linha ou Incondicional.

**Dimensões.** Permite especificar a dimensionalidade de uma ou mais soluções de ajuste de escala. Uma solução é calculada para cada número no intervalo. Especifique números inteiros entre 1 e 6; um mínimo de 1 é permitido apenas se você selecionar **Distância euclidiana** como o modelo de ajuste de escala. Para uma única solução, especifique o mesmo número para mínimo e máximo.

**Modelo de Ajuste de Escala.** Permite especificar as suposições pelas quais o ajuste de escala é executado. As alternativas disponíveis são Distância euclidiana ou Diferenças individuais Distância euclidiana (também conhecida como INDSCAL). Para o modelo de distância euclidiana de diferenças individuais, é possível selecionar **Permitir ponderações de assunto negativas**, se apropriado para seus dados.

## Opções de ajuste de escala multidimensional

É possível especificar opções para sua análise de ajuste de escala multidimensional.

**Exibição.** Permite selecionar vários tipos de saída. As opções disponíveis são gráficos de Grupo, gráficos das observações individuais, Matriz de dados e Sumarização de modelo e opções.

**Crítérios.** Permite determinar quando a iteração deve parar. Para alterar os padrões, insira valores para **Convergência de estresse de S**, **Valor mínimo de estresse de S** e **Máximo de iterações**.

**Tratar distâncias menores que n como omissas.** Distâncias que forem menores que este valor são excluídas da análise.

## Recursos adicionais do comando ALSCAL

O idioma da sintaxe de comando também permite:

- Utilizar três tipos de modelo adicionais, conhecidos como ASCAL, AINDS e GEMSCAL, na literatura sobre ajuste de escala multidimensional.
- Executar transformações polinomiais nos dados de intervalo e de proporções.
- Analisar similaridades (ao invés de distâncias) com dados ordinais.
- Analisar dados nominais.
- Salvar várias matrizes de coordenadas e de ponderações em arquivos e lê-los novamente para análise.
- Restringir desdobra multidimensional.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

## Estatísticas de razão

---

O procedimento Estatísticas de Razão fornece uma lista abrangente de estatísticas básicas para descrever a razão entre duas variáveis de escala.

É possível classificar a saída por valores de uma variável de agrupamento em ordem crescente ou decrescente. O relatório de estatísticas de razão pode ser suprimido na saída e os resultados podem ser salvos em um arquivo externo.

### Exemplo

Existe uma boa uniformidade na relação entre o preço de avaliação e o preço de venda das casas em cada um dos cinco municípios? A partir da saída, você pode aprender que a distribuição das razões varia consideravelmente de município para município.

### Estatísticas

Mediana, média, média ponderada, intervalos de confiança, coeficiente de dispersão (COD), coeficiente de variação (COV), coeficiente de variação centrado na mediana, coeficiente de variação centrado na média, viés relacionado ao preço (PRB), diferencial relacionado ao preço (PRD), desvio padrão, desvio médio absoluto (AAD), intervalo, valores mínimo e máximo e o índice de concentração calculado para um intervalo ou porcentagem especificada pelo usuário dentro da razão mediana.

## Considerações de dados

### Dados

Utilize códigos ou sequências de caracteres numéricos para codificar variáveis de agrupamento (níveis de medição nominais ou ordinais).

### Suposições

As variáveis que definem o numerador e o denominador da razão devem ser variáveis de escala que assumem valores positivos.

## Obtendo estatísticas de razão

1. A partir dos menus, escolha:

## **Analisar > Estatísticas Descritivas > Razão...**

2. Selecione uma variável numeradora.
3. Selecione uma variável de denominador.

Opcionalmente:

- Selecione uma variável de agrupamento e especifique a ordem dos grupos nos resultados.
- Escolha se deseja exibir os resultados no visualizador.
- Escolha incluir uma palavra-chave **N** no subcomando **PRINT**. A configuração inclui os tamanhos da amostra na tabela de estatísticas na saída.
- Escolha se deseja salvar os resultados em um arquivo externo para uso posterior, e especifique o nome do arquivo no qual os resultados são salvos.

## **Estatísticas de razão: estatísticas**

### **Tendência central**

Medidas de tendência central são estatísticas que descrevem a distribuição de razões.

#### **Mediana**

O valor tal que o número de razões que são menores que este valor e o número de razões que são maiores que este valor são os mesmos.

#### **Média**

O resultado da soma de razões e da divisão do resultado pelo número total de razões.

#### **Média ponderada**

O resultado da divisão da média do numerador pela média do denominador. A média ponderada também é a média das razões ponderadas pelo denominador.

### **Intervalos de confiança para o PRB e medidas de tendência central**

Exibe intervalos de confiança para PRB, a média, a mediana e a média ponderada (se solicitada).

Especifique um valor maior ou igual a 0 e menor que 100 como o nível de confiança. A configuração padrão é 95%.

### **Dispersão**

Essas estatísticas medem a quantidade de variação, ou dispersão, nos valores observados.

#### **AAD**

O desvio absoluto médio é o resultado da soma dos desvios absolutos das razões em torno da mediana e da divisão do resultado pelo número total de razões.

#### **COD**

O coeficiente de dispersão é o resultado da expressão do desvio absoluto médio como percentual da mediana.

#### **COV**

O coeficiente de variação. O coeficiente de variação centrado na mediana é o resultado da expressão da raiz quadrada média do desvio da mediana como uma porcentagem da mediana. O coeficiente de variação centrado na média é o resultado da expressão do desvio padrão como uma porcentagem da média.

#### **PRB**

O viés relacionado ao preço. Um índice para determinar se as relações entre avaliação e preço são sistematicamente mais altas ou mais baixas para propriedades com preços mais altos. Regride as diferenças percentuais nos índices de avaliação do índice mediano na base 2 logaritmos das medidas de proxy de valor calculadas como as médias dos preços de venda e os índices dos valores avaliados no índice mediano. Fornece a alteração percentual nas taxas de avaliação para uma alteração de 100% no valor.

#### **PRD**

O diferencial relacionado ao preço, também conhecido como índice de regressividade, é o resultado da divisão da média pela média ponderada.

**Desvio padrão**

O desvio padrão é o resultado da soma dos desvios quadrados das razões em torno da média, dividindo o resultado pelo número total de razões menos um e tirando a raiz quadrada positiva.

**Intervalo**

O intervalo é o resultado da subtração da razão mínima da razão máxima.

**Mínimo**

O mínimo é a menor razão.

**Máximo**

O máximo é a maior razão.

**Índice de concentração**

O coeficiente de concentração mede a porcentagem de razões que caem em um intervalo. Ele pode ser calculado de duas maneiras diferentes:

**Entre proporções**

Aqui o intervalo é definido explicitamente especificando os valores baixo e alto do intervalo. Insira os valores para a proporção baixa e proporção alta e clique em **Incluir** para obter um intervalo.

**Dentro da porcentagem da mediana**

Aqui o intervalo é definido implicitamente especificando a porcentagem da mediana. Insira um valor entre 0 e 100 e clique em **Incluir**. A extremidade inferior do intervalo é igual a  $(1 - 0,01 \times \text{valor}) \times \text{mediana}$ , e a extremidade superior é igual a  $(1 + 0,01 \times \text{valor}) \times \text{mediana}$ .

## Estatísticas de razão

---

O procedimento Estatísticas de Razão fornece uma lista abrangente de estatísticas básicas para descrever a razão entre duas variáveis de escala.

É possível classificar a saída por valores de uma variável de agrupamento em ordem crescente ou decrescente. O relatório de estatísticas de razão pode ser suprimido na saída e os resultados podem ser salvos em um arquivo externo.

**Exemplo**

Existe uma boa uniformidade na relação entre o preço de avaliação e o preço de venda das casas em cada um dos cinco municípios? A partir da saída, você pode aprender que a distribuição das razões varia consideravelmente de município para município.

**Estatísticas**

Mediana, média, média ponderada, intervalos de confiança, coeficiente de dispersão (COD), coeficiente de variação (COV), coeficiente de variação centrado na mediana, coeficiente de variação centrado na média, viés relacionado ao preço (PRB), diferencial relacionado ao preço (PRD), desvio padrão, desvio médio absoluto (AAD), intervalo, valores mínimo e máximo e o índice de concentração calculado para um intervalo ou porcentagem especificada pelo usuário dentro da razão mediana.

**Considerações de dados****Dados**

Utilize códigos ou sequências de caracteres numéricos para codificar variáveis de agrupamento (níveis de medição nominais ou ordinais).

**Suposições**

As variáveis que definem o numerador e o denominador da razão devem ser variáveis de escala que assumem valores positivos.

**Obtendo estatísticas de razão**

1. A partir dos menus, escolha:

**Analisar > Estatísticas Descritivas > Razão...**

2. Selecione uma variável numeradora.

3. Selecione uma variável de denominador.

Opcionalmente:

- Selecione uma variável de agrupamento e especifique a ordem dos grupos nos resultados.
- Escolha se deseja exibir os resultados no visualizador.
- Escolha incluir uma palavra-chave **N** no subcomando **PRINT**. A configuração inclui os tamanhos da amostra na tabela de estatísticas na saída.
- Escolha se deseja salvar os resultados em um arquivo externo para uso posterior, e especifique o nome do arquivo no qual os resultados são salvos.

## Estatísticas de razão: estatísticas

### Tendência central

Medidas de tendência central são estatísticas que descrevem a distribuição de razões.

#### Mediana

O valor tal que o número de razões que são menores que este valor e o número de razões que são maiores que este valor são os mesmos.

#### Média

O resultado da soma de razões e da divisão do resultado pelo número total de razões.

#### Média ponderada

O resultado da divisão da média do numerador pela média do denominador. A média ponderada também é a média das razões ponderadas pelo denominador.

### Intervalos de confiança para o PRB e medidas de tendência central

Exibe intervalos de confiança para PRB, a média, a mediana e a média ponderada (se solicitada). Especifique um valor maior ou igual a 0 e menor que 100 como o nível de confiança. A configuração padrão é 95%.

### Dispersão

Essas estatísticas medem a quantidade de variação, ou dispersão, nos valores observados.

#### AAD

O desvio absoluto médio é o resultado da soma dos desvios absolutos das razões em torno da mediana e da divisão do resultado pelo número total de razões.

#### COD

O coeficiente de dispersão é o resultado da expressão do desvio absoluto médio como percentual da mediana.

#### COV

O coeficiente de variação. O coeficiente de variação centrado na mediana é o resultado da expressão da raiz quadrada média do desvio da mediana como uma porcentagem da mediana. O coeficiente de variação centrado na média é o resultado da expressão do desvio padrão como uma porcentagem da média.

#### PRB

O viés relacionado ao preço. Um índice para determinar se as relações entre avaliação e preço são sistematicamente mais altas ou mais baixas para propriedades com preços mais altos. Regride as diferenças percentuais nos índices de avaliação do índice mediano na base 2 logaritmos das medidas de proxy de valor calculadas como as médias dos preços de venda e os índices dos valores avaliados no índice mediano. Fornece a alteração percentual nas taxas de avaliação para uma alteração de 100% no valor.

#### PRD

O diferencial relacionado ao preço, também conhecido como índice de regressividade, é o resultado da divisão da média pela média ponderada.

#### Desvio padrão

O desvio padrão é o resultado da soma dos desvios quadrados das razões em torno da média, dividindo o resultado pelo número total de razões menos um e tirando a raiz quadrada positiva.

#### Intervalo

O intervalo é o resultado da subtração da razão mínima da razão máxima.

**Mínimo**

O mínimo é a menor razão.

**Máximo**

O máximo é a maior razão.

**Índice de concentração**

O coeficiente de concentração mede a porcentagem de razões que caem em um intervalo. Ele pode ser calculado de duas maneiras diferentes:

**Entre proporções**

Aqui o intervalo é definido explicitamente especificando os valores baixo e alto do intervalo. Insira os valores para a proporção baixa e proporção alta e clique em **Incluir** para obter um intervalo.

**Dentro da porcentagem da mediana**

Aqui o intervalo é definido implicitamente especificando a porcentagem da mediana. Insira um valor entre 0 e 100 e clique em **Incluir**. A extremidade inferior do intervalo é igual a  $(1 - 0,01 \times \text{valor}) \times \text{mediana}$ , e a extremidade superior é igual a  $(1 + 0,01 \times \text{valor}) \times \text{mediana}$ .

## Gráfico P-P

---

O procedimento de gráficos P-P produz gráficos de probabilidade de uma ou mais variáveis de sequência ou série temporal. As variáveis podem ser padronizadas, diferenciadas e transformadas antes da plotagem.

As distribuições de teste disponíveis incluem beta, qui-quadrado, exponencial, gama, semi-normal, Laplace, logístico, Log-normal, normal, Pareto, t, Weibull e uniforme. Dependendo da distribuição selecionada, é possível especificar graus de liberdade e outros parâmetros.

- Você obtém gráficos de probabilidade para valores transformados. As opções de transformação incluem log natural, padronização de valores, diferença e diferença sazonal.
- É possível especificar o método para calcular as distribuições esperadas e para resolver "empates" (múltiplas observações com o mesmo valor).

**Distribuição do teste**

Especifique um tipo de distribuição para seus dados. A lista suspensa fornece as opções a seguir:

**Beta**

*Distribuição beta.* Os parâmetros shape1 e shape2  $a$  e  $b$  devem ser positivos. Se eles não forem especificados, DISTRIBUTION os estimará a partir da média amostral e do desvio padrão amostral. Todas as observações devem estar no intervalo de 0 a 1, inclusive.

**Qui-quadrado**

*Distribuição de qui-quadrado.* Deve-se especificar os graus de liberdade ( $df$ ). Não são permitidas observações negativas.

**Exponencial**

*Distribuição exponencial.* O parâmetro de escala  $a$  deve ser positivo. Se o parâmetro não for especificado, DISTRIBUTION o estima a partir da média da amostra. Não são permitidas observações negativas.

**Gama**

*Distribuição de Gamma.* Os parâmetros de forma e escala  $a$  e  $b$  devem ser positivos. Se eles não forem especificados, DISTRIBUTION os estimará a partir da média amostral e do desvio padrão amostral. Não são permitidas observações negativas.

**Meio-normal**

*Distribuição semi-normal.* Os dados são considerados livres de localização ou centralizados. (Parâmetro de localização=0.) É possível especificar o parâmetro de escala  $a$  ou deixar DISTRIBUTION estimá-lo usando o método de máxima verossimilhança.

**Laplace**

*Laplace ou distribuição exponencial dupla.* LAPLACE recebe uma localização e um parâmetro de escala ( $a$  e  $b$ ). O parâmetro de escala ( $b$ ) deve ser positivo. Se os parâmetros não forem especificados, DISTRIBUTION os estima a partir da média amostral e do desvio padrão amostral.

**Logística**

*Distribuição logística.* LOGISTIC recebe uma localização e um parâmetro de escala ( $a$  e  $b$ ). O parâmetro de escala ( $b$ ) deve ser positivo. Se os parâmetros não forem especificados, DISTRIBUTION os estima a partir da média amostral e do desvio padrão amostral.

**Log-normal**

*Distribuição logística.* Os parâmetros de escala e forma  $a$  e  $b$  devem ser positivos. Se eles não forem especificados, DISTRIBUTION os estima a partir da média e do desvio padrão do logaritmo natural dos dados da amostra. Não são permitidas observações negativas.

**Normal**

*Distribuição normal.* O parâmetro de localização  $a$  pode ser qualquer valor numérico, enquanto o parâmetro de escala  $b$  deve ser positivo. Se eles não forem especificados, DISTRIBUTION os estimará a partir da média amostral e do desvio padrão amostral.

**Pareto**

*Distribuição de pareto.* Os parâmetros limite e de forma  $a$  e  $b$  devem ser positivos. Se não forem especificados, DISTRIBUTION assume que  $a$  é igual à observação mínima e estima  $b$  pelo método da máxima verossimilhança. Não são permitidas observações negativas.

**T de Student**

*Distribuição t de aluno.* Deve-se especificar os graus de liberdade ( $df$ ).

**Uniforme**

*Distribuição uniforme.* UNIFORM recebe um parâmetro mínimo e máximo ( $a$  e  $b$ ). O parâmetro  $a$  deve ser igual ou maior que  $b$ . Se os parâmetros não forem especificados, DISTRIBUTION os assume a partir dos dados de amostra.

**Weibull**

*Distribuição de Weibull.* Os parâmetros de escala e forma  $a$  e  $b$  devem ser positivos. Se eles não forem especificados, DISTRIBUTION os estima usando o método dos mínimos quadrados. Não são permitidas observações negativas.

**Parâmetros de distribuição**

Fornece estratégia de distribuição e opções de parâmetros.

**Estimar a partir dos dados**

Quando selecionada, esta configuração estima os parâmetros de distribuição com base nos dados e no tipo de distribuição selecionado.

**Especificar**

Quando selecionado, é possível especificar os parâmetros de distribuição para o tipo de distribuição selecionado.

**Nota:** Os parâmetros disponíveis variam de acordo com o tipo de distribuição selecionado.

**Transformar**

As opções fornecidas definem as configurações de transformação e periodicidade.

**Transformação do log natural**

Transforma os dados usando o logaritmo natural (base  $e$ ) para remover amplitude variável.

**Valores padronizados**

Transforma as variáveis de sequência ou série temporal em uma amostra com média 0 e desvio padrão 1.

**Diferença**

Especifica o grau de diferenciação que é usado antes da plotagem para converter uma variável não estacionária em uma variável estacionária com média e variância constantes. Insira um valor apropriado no campo.

**Diferença sazonal**

Se a variável exibir um padrão sazonal ou periódico, é possível usar essa configuração para diferenciar sazonalmente a variável antes de plotar.

**Nota:** Essa configuração é habilitada somente quando uma variável de sequência ou série temporal, com periodicidade definida, é selecionada como uma das variáveis quantitativas.



### Fórmula de estimação de proporção

As opções fornecidas definem a fórmula usada para estimar proporções.

#### de Blom

Cria uma nova variável de classificação com base nas estimativas de proporção que usam a fórmula  $(r-3/8) / (w+1/4)$ , em que  $w$  é a soma dos pesos dos casos e  $r$  é a classificação.

#### Rankit

Usa a fórmula  $(r-1/2) / w$ , em que  $w$  é o número de observações e  $r$  é a classificação, variando de 1 a  $w$ .

#### de Tukey

Usa a fórmula  $(r-1/3) / (w+1/3)$ , em que  $r$  é a classificação e  $w$  é a soma dos pesos de caso.

#### de Van der Waerden

A transformação de Van der Waerden, definida pela fórmula  $r / (w+1)$ , em que  $w$  é a soma dos pesos dos casos e  $r$  é a classificação, variando de 1 a  $w$ .

### Classificação atribuída aos empates

As opções fornecidas controlam o método para determinar como lidar com valores de empate. A tabela a seguir mostra como os diferentes métodos atribuem classificações a valores vinculados.

Tabela 3. Métodos e resultados de classificação

Valor	Média	Baixa	Alta	Quebrar vínculos arbitrariamente
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

### Obtendo gráficos de probabilidade P-P

1. A partir dos menus, escolha:

**Analisar > Estatísticas descritivas > Gráficos P-P...**

2. Selecione uma ou mais variáveis numéricas e mova-as para o campo Variáveis.

3. Selecione uma distribuição de teste.

Opcionalmente, é possível selecionar opções de transformação para obter gráficos de probabilidade para valores transformados e especificar um método para calcular as distribuições esperadas.

## Gráfico Q-Q

O procedimento de gráficos Q-Q produz gráficos de probabilidade para valores transformados. As distribuições de teste disponíveis incluem beta, qui-quadrado, exponencial, gama, semi-normal, Laplace, logístico, Log-normal, normal, Pareto, t, Weibull e uniforme. Dependendo da distribuição selecionada, é possível especificar graus de liberdade e outros parâmetros.

- Você obtém gráficos de probabilidade para valores transformados. As opções de transformação incluem log natural, padronização de valores, diferença e diferença sazonal.
- É possível especificar o método para calcular as distribuições esperadas e para resolver "empates" (múltiplas observações com o mesmo valor).

#### Distribuição do teste

Especifique um tipo de distribuição para seus dados. A lista suspensa fornece as opções a seguir:

**Beta**

*Distribuição beta.* Os parâmetros shape1 e shape2  $a$  e  $b$  devem ser positivos. Se eles não forem especificados, DISTRIBUTION os estimará a partir da média amostral e do desvio padrão amostral. Todas as observações devem estar no intervalo de 0 a 1, inclusive.

**Qui-quadrado**

*Distribuição de qui-quadrado.* Deve-se especificar os graus de liberdade ( $df$ ). Não são permitidas observações negativas.

**Exponencial**

*Distribuição exponencial.* O parâmetro de escala  $a$  deve ser positivo. Se o parâmetro não for especificado, DISTRIBUTION o estima a partir da média da amostra. Não são permitidas observações negativas.

**Gama**

*Distribuição de Gamma.* Os parâmetros de forma e escala  $a$  e  $b$  devem ser positivos. Se eles não forem especificados, DISTRIBUTION os estimará a partir da média amostral e do desvio padrão amostral. Não são permitidas observações negativas.

**Meio-normal**

*Distribuição semi-normal.* Os dados são considerados livres de localização ou centralizados. (Parâmetro de localização=0.) É possível especificar o parâmetro de escala  $a$  ou deixar DISTRIBUTION estimá-lo usando o método de máxima verossimilhança.

**Laplace**

*Laplace ou distribuição exponencial dupla.* LAPLACE recebe uma localização e um parâmetro de escala ( $a$  e  $b$ ). O parâmetro de escala ( $b$ ) deve ser positivo. Se os parâmetros não forem especificados, DISTRIBUTION os estima a partir da média amostral e do desvio padrão amostral.

**Logística**

*Distribuição logística.* LOGISTIC recebe uma localização e um parâmetro de escala ( $a$  e  $b$ ). O parâmetro de escala ( $b$ ) deve ser positivo. Se os parâmetros não forem especificados, DISTRIBUTION os estima a partir da média amostral e do desvio padrão amostral.

**Log-normal**

*Distribuição logística.* Os parâmetros de escala e forma  $a$  e  $b$  devem ser positivos. Se eles não forem especificados, DISTRIBUTION os estima a partir da média e do desvio padrão do logaritmo natural dos dados da amostra. Não são permitidas observações negativas.

**Normal**

*Distribuição normal.* O parâmetro de localização  $a$  pode ser qualquer valor numérico, enquanto o parâmetro de escala  $b$  deve ser positivo. Se eles não forem especificados, DISTRIBUTION os estimará a partir da média amostral e do desvio padrão amostral.

**Pareto**

*Distribuição de pareto.* Os parâmetros limite e de forma  $a$  e  $b$  devem ser positivos. Se não forem especificados, DISTRIBUTION assume que  $a$  é igual à observação mínima e estima  $b$  pelo método da máxima verossimilhança. Não são permitidas observações negativas.

**T de Student**

*Distribuição t de aluno.* Deve-se especificar os graus de liberdade ( $df$ ).

**Uniforme**

*Distribuição uniforme.* UNIFORM recebe um parâmetro mínimo e máximo ( $a$  e  $b$ ). O parâmetro  $a$  deve ser igual ou maior que  $b$ . Se os parâmetros não forem especificados, DISTRIBUTION os assume a partir dos dados de amostra.

**Weibull**

*Distribuição de Weibull.* Os parâmetros de escala e forma  $a$  e  $b$  devem ser positivos. Se eles não forem especificados, DISTRIBUTION os estima usando o método dos mínimos quadrados. Não são permitidas observações negativas.

**Parâmetros de distribuição**

Fornece estratégia de distribuição e opções de parâmetros.

**Estimar a partir dos dados**

Quando selecionada, esta configuração estima os parâmetros de distribuição com base nos dados e no tipo de distribuição selecionado.

**Especificar**

Quando selecionado, é possível especificar os parâmetros de distribuição para o tipo de distribuição selecionado.

**Nota:** Os parâmetros disponíveis variam de acordo com o tipo de distribuição selecionado.

**Transformar**

As opções fornecidas definem as configurações de transformação e periodicidade.

**Transformação do log natural**

Transforma os dados usando o logaritmo natural (base  $e$ ) para remover amplitude variável.

**Valores padronizados**

Transforma as variáveis de sequência ou série temporal em uma amostra com média 0 e desvio padrão 1.

**Diferença**

Especifica o grau de diferenciação que é usado antes da plotagem para converter uma variável não estacionária em uma variável estacionária com média e variância constantes. Insira um valor apropriado no campo.

**Diferença sazonal**

Se a variável exibir um padrão sazonal ou periódico, é possível usar essa configuração para diferenciar sazonalmente a variável antes de plotar.

**Nota:** Essa configuração é habilitada somente quando uma variável de sequência ou série temporal, com periodicidade definida, é selecionada como uma das variáveis quantitativas.

**Fórmula de estimação de proporção**

As opções fornecidas definem a fórmula usada para estimar proporções.

**de Blom**

Cria uma nova variável de classificação com base nas estimativas de proporção que usam a fórmula  $(r-3/8) / (w+1/4)$ , em que  $w$  é a soma dos pesos dos casos e  $r$  é a classificação.

**Rankit**

Usa a fórmula  $(r-1/2) / w$ , em que  $w$  é o número de observações e  $r$  é a classificação, variando de 1 a  $w$ .

**de Tukey**

Usa a fórmula  $(r-1/3) / (w+1/3)$ , em que  $r$  é a classificação e  $w$  é a soma dos pesos de caso.

**de Van der Waerden**

A transformação de Van der Waerden, definida pela fórmula  $r/(w+1)$ , em que  $w$  é a soma dos pesos dos casos e  $r$  é a classificação, variando de 1 a  $w$ .

**Classificação atribuída aos empates**

As opções fornecidas controlam o método para determinar como lidar com valores de empate. A tabela a seguir mostra como os diferentes métodos atribuem classificações a valores vinculados.

*Tabela 4. Métodos e resultados de classificação*

Valor	Média	Baixa	Alta	Quebrar vínculos arbitrariamente
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3

Tabela 4. Métodos e resultados de classificação (continuação)

Valor	Média	Baixa	Alta	Quebrar vínculos arbitrariamente
20	6	6	6	4

## Obtendo gráficos de probabilidade Q-Q

1. A partir dos menus, escolha:

**Analisar > Estatísticas descritivas > Gráficos Q-Q...**

2. Selecione uma ou mais variáveis numéricas e mova-as para o campo Variáveis.

3. Selecione uma distribuição de teste.

Opcionalmente, é possível selecionar opções de transformação para obter gráficos de probabilidade para valores transformados e especificar um método para calcular as distribuições esperadas.

## Análise de Característica de Operação do Receptor

A Análise de Característica de operação do receptor (ROC) é uma maneira útil de avaliar a precisão de predições do modelo, plotando a sensibilidade versus (1-especificidade) de um teste de classificação (conforme o limite varia em uma gama inteira de resultados de teste de diagnóstico). A área completa sob uma determinada curva ROC, ou AUC, formula uma estatística importante que representa a probabilidade de que a predição estará na ordem correta quando uma variável de teste for observada (para um sujeito que é selecionado aleatoriamente do grupo de casos e o outro selecionado aleatoriamente do grupo de controle). A Análise de ROC suporta a inferência relacionada a curvas de um único AUC, de precisão-rechamada (PR) e fornece opções para comparar duas curvas ROC que são geradas a partir de grupos independentes ou de assuntos pareados.

O antigo procedimento de Curva ROC suporta a inferência estatística sobre uma única curva ROC. Isso também pode ser recuperado pelo novo procedimento de Análise de ROC. Além disso, o novo procedimento de Análise de ROC pode comparar duas curvas ROC geradas a partir de grupos independentes ou de assuntos empardos.

As curvas PR plotam precisão versus rechamada, tendem a ser mais informativas quando as amostras de dados observadas estão altamente defasadas e fornecem uma alternativa a curvas ROC para dados com uma grande defasagem na distribuição de classes.

### Exemplo:

É de interesse de um banco classificar corretamente clientes nos clientes que serão e que não serão inadimplentes em seus empréstimos; portanto, são desenvolvidos modelos especiais para tomar essas decisões. A Análise de ROC pode ser usada para avaliar a precisão das predições do modelo.

### Estatísticas

AUC, grupo negativo, valores omissos, classificação positiva, valor de corte, força de convicção, intervalo de confiança assintótica bilateral, distribuição, erro padrão, design de grupo independente, design de amostra pareado, suposição não paramétrica, suposição de distribuição exponencial binegativa, ponto médio, ponto de corte, curva PR, interpolação stepwise, significância assintótica (bicaudal), Sensibilidade e (1-Especificidade), Precisão e Rechamada.

### Métodos

São comparadas as áreas sob duas curvas ROC, que são geradas de grupos independentes ou de assuntos pareados. A comparação de duas curvas ROC pode fornecer informações adicionais na precisão resultante de duas abordagens de diagnóstico comparativas.

## Considerações sobre dados de Análise de ROC

### Dados

As curvas PR plotam Precisão versus Rechamada e tendem a ser mais informativas quando as amostras de dados observadas estão altamente defasadas. Uma interpolação linear simples pode gerar erroneamente uma estimativa excessivamente otimista de uma curva PR.

### Suposições

A predição estará na ordem correta quando uma variável de teste for observada para um assunto selecionado aleatoriamente do grupo de caso e o outro selecionado aleatoriamente do grupo de controle. Cada grupo definido conterà pelo menos uma observação válida. Somente uma única variável de agrupamento é usada para um único procedimento.

## Obtendo uma Análise de ROC

1. Nos menus, escolha:

**Analisar > Classificar > Análise de ROC**

2. Selecione uma ou mais variáveis de probabilidade de teste.

3. Selecione uma variável de estado.

4. Identifique o valor de *positivo* para a variável de estado.

5. Opcionalmente, selecione a opção **Design de amostra pareado** ou selecione uma única variável de agrupamento (não é possível selecionar ambas as opções).

- Use a configuração **Design de amostra pareado** para solicitar o design de amostra pareado para a variável ou variáveis de teste. O design de amostra pareado compara duas curvas ROC em um cenário de amostra pareado quando diversos valores de teste são medidos nos mesmos assuntos que estão associados a uma variável de estado.

**Nota:** Quando **Design de amostra pareado** é selecionado, as opções **Variável de agrupamento** e **Suposição de distribuição** (no diálogo Opções) são desativadas.

- Quando uma variável de agrupamento numérico é selecionada, você pode clicar em **Definir Grupos ...** para solicitar o design de grupo independente para a (s) variável (s) de teste (s), e para especificar dois valores, um ponto intermediário ou um ponto de corte.

6. Opcionalmente, clique em Opções para definir as configurações de classificação, de direção de teste, de parâmetros de erro padrão e de valores omissos.

7. Opcionalmente, clique em Exibir para definir as configurações de plotagem e impressão (que incluem as configurações da Curva ROC, da Curva de Precisão-Rechamada e de qualidade de modelo).

8. Clique em **OK**.

## Análise ROC: opções

É possível especificar as seguintes opções para sua análise ROC:

### Classificação

Permite especificar se o valor de corte deve ser incluído ou excluído ao criar uma classificação *positiva*. Esta configuração atualmente não tem nenhum efeito na saída.

### Direção do teste

Fornece opções para especificar qual direção da variável de resultado de teste indica aumento de força de convicção de que o assunto é teste positivo.

### Parâmetros para o erro padrão da área

Permite especificar o método para estimar o erro padrão da área abaixo da curva. Os métodos disponíveis são não paramétricos e exponenciais binegativos. A configuração padrão **Não paramétrico** fornece estimativas sob a suposição não paramétrica. A configuração **Exponencial binegativa** fornece estimativas sob a suposição de distribuição exponencial binegativa.

A seção também permite especificar o nível de confiança para o intervalo de confiança assintótico bilateral do AUC. O intervalo disponível é de 0,0% a 100,0% (o valor padrão é 95%).

**Nota:** A configuração se aplica somente ao design de grupo independente e não tem efeito no design de amostra pareada.

### Valores omissos

Permite especificar como os valores ausentes são tratados. Quando a configuração não é selecionada, os valores omissos do usuário e os valores omissos do sistema são excluídos. Quando a configuração é selecionada, os valores omissos do usuário são tratados como válidos e os valores omissos do sistema são excluídos. Os casos com valores omissos do sistema, na variável de teste ou na variável de estado, são sempre excluídos da análise.

## Análise de Característica de Operação do Receptor: Exibição

É possível especificar as seguintes configurações de exibição para sua análise de Característica de Operação do Receptor:

### Gráfico

Fornecer opções para plotar as curvas ROC e de Precisão-Rechamada.

#### Curva ROC

Quando selecionada, o gráfico Curva ROC é exibido na saída. Selecione **Com linha de referência diagonal** para desenhar uma linha de referência diagonal no gráfico de Curva ROC.

#### Curva de Precisão-Rechamada

Quando selecionada, o gráfico de Curva de Precisão-Rechamada é exibido na saída. As Curvas de Precisão-Rechamada tendem a ser mais informativas quando as amostras de dados observadas estão altamente defasadas e fornecem uma alternativa às Curvas ROC para dados com uma grande defasagem na distribuição de classe. A configuração padrão **Interpolar nos positivos verdadeiros** faz a interpolação stepwise nos positivos verdadeiros. A configuração **Interpolar nos falsos positivos** faz a interpolação stepwise nos falsos positivos.

#### Qualidade modelo global

A configuração controla se um gráfico de barras é criado ou não para exibir o valor do limite inferior do intervalo de confiança do AUC estimado. Por padrão, a configuração não é selecionada, o que suprime o gráfico de barras.

### Imprimir

Fornecer opções para definir a saída para as estatísticas correspondentes.

#### Erro padrão e intervalo de confiança

A configuração controla quais estatísticas são exibidas na tabela "Área sob a curva". Quando a configuração não é selecionada, somente o AUC estimado é exibido. Quando a configuração é selecionada, as estatísticas adicionais são exibidas, incluindo o erro padrão do AUC, a significância assintótica (bicaudal) e os limites de intervalo de confiança assintótico sob a hipótese nula.

#### Pontos de coordenadas da curva ROC

A configuração controla os pontos de coordenadas da curva ROC, junto com os valores de corte. Quando a configuração não é selecionada, a saída dos pontos de coordenadas é suprimida. Quando a configuração for selecionada, os pares de valores de Sensibilidade e (1-Especificidade) serão fornecidos com os valores de corte para cada curva ROC.

#### Índice do Youden

Exibe o valor do Índice de Youden para cada valor de corte da curva ROC.

#### Pontos de coordenadas da curva de Precisão-Rechamada

A configuração controla os pontos de coordenadas da curva de Precisão-Rechamada, junto com os valores de corte. Quando a configuração não é selecionada, a saída dos pontos de coordenadas é suprimida. Quando a configuração for selecionada, os pares de valores de Precisão e Rechamada serão fornecidos com os valores de corte para cada curva de Precisão-Rechamada.

#### Métricas de avaliação do classificador

A configuração controla a exibição da tabela Métricas de avaliação do classificador na saída. A tabela mostra como um modelo de classificação ajusta os dados em comparação com uma designação aleatória e fornece as seguintes informações:

- As variáveis de teste especificadas pelo usuário
- Informações sobre o grupo
- Índice de Gini (o índice de Gini é  $2 \cdot \text{AUC} - 1$ , em que AUC é a área sob a curva ROC)
- Valores máximos de K-S e de corte

## Análise de ROC: Definir grupos (sequência)

Para variáveis de agrupamento de sequências de caracteres, insira uma sequência de caracteres para o Grupo 1 e outro valor para o Grupo 2, como *yes* e *no*. Casos com outras sequências de caracteres são excluídos da análise.

**Nota:** Os valores especificados devem existir na variável, caso contrário, será exibida uma mensagem de erro, indicando que pelo menos um dos grupos está vazio.

## Análise de ROC: Definir grupos (numéricos)

Para variáveis de agrupamento numéricas, defina os dois grupos para o teste *t* especificando dois valores, um ponto médio ou um ponto de corte.

**Nota:** Os valores especificados devem existir na variável, caso contrário, será exibida uma mensagem de erro, indicando que pelo menos um dos grupos está vazio.

- **Utilizar valores especificados.** Insira um valor para o Grupo 1 e outro valor para o Grupo 2. Casos com quaisquer outros valores são excluídos da análise. Os números não precisam ser inteiros (por exemplo, 6,25 e 12,5 são válidos).
- **Usar valor de ponto médio.** Quando selecionado, os grupos são separados em valores de ponto médio  $<$  e  $\geq$ .
- **Use ponto de corte.**
  - **Ponto de Corte.** Insira um número que divide os valores da variável de agrupamento em dois conjuntos. Todos os casos com valores menores que o ponto de corte formam um grupo e os casos com valores maiores ou iguais ao ponto de corte formam o outro grupo.

## Curvas ROC

---

Este procedimento é uma maneira útil para avaliar o desempenho de esquemas de classificação em que existe uma variável com duas categorias pelas quais os assuntos são classificados.

**Exemplo.** Um banco deseja classificar corretamente quais clientes irão ou não ficar inadimplentes com seus empréstimos, de modo que métodos especiais são desenvolvidos para tomar essas decisões. As curvas ROC podem ser utilizadas para avaliar o grau de desempenho destes métodos.

**Estatísticas.** Área sob a curva ROC com um intervalo de confiança e pontos de coordenadas da curva ROC. Gráficos: curva ROC.

**Métodos.** A estimativa da área sob a curva ROC pode ser calculada de modo paramétrico ou não paramétrico usando um modelo exponencial binomial.

Considerações de Dados da Curva ROC

O antigo procedimento de Curva ROC suporta a inferência estatística sobre uma única curva ROC. Isso também pode ser recuperado pelo novo procedimento de Análise de ROC. Além disso, o novo procedimento de Análise de ROC pode comparar duas curvas ROC geradas a partir de grupos independentes ou de assuntos empardos.

**Dados.** As variáveis de teste são quantitativas. As variáveis de teste são geralmente compostas de probabilidades de análise discriminante ou de regressão logística, ou compostas de escores em uma escala arbitrária indicando a "intensidade de convicção" de um avaliador de que um assunto se enquadre em uma ou outra categoria. A variável de estado pode ser de qualquer tipo e indica a categoria real à qual um assunto pertence. O valor da variável de estado indica qual categoria deve ser considerada *positiva*.

**Suposições.** Supõe-se que números maiores na escala do avaliador representam uma crença maior de que o assunto pertence a uma categoria, ao passo que números menores na escala representam uma crença maior de que o assunto pertence à outra categoria. O usuário deve escolher qual direção é *positiva*. Também supõe-se que a categoria *real* à qual cada assunto pertence é conhecida.

Para Obter uma Curva ROC

1. Nos menus, escolha:  
**Analisar > Classificar > Curva ROC...**
2. Selecione uma ou mais variáveis de probabilidade de teste.
3. Selecione uma variável de estado.
4. Identifique o valor de *positivo* para a variável de estado.

## Opções de curva ROC

É possível especificar as opções a seguir para sua análise ROC:

### Classificação

Permite especificar se o valor de corte deve ser incluído ou excluído ao criar uma classificação *positiva*. Atualmente, esta configuração não tem efeito na saída.

### Direção de teste

Permite especificar a direção da escala com relação à categoria *positiva*.

### Parâmetros para erro padrão da área

Permite especificar o método para estimar o erro padrão da área abaixo da curva. Os métodos disponíveis são exponenciais não paramétricos e binegativos. Também permite configurar o nível do intervalo de confiança. O intervalo disponível é de 50,1% a 99,9%.

### Valores omissos

Permite especificar como os valores ausentes são tratados.

## simulação

---

Modelos preditivos, como regressão linear, requerem um conjunto de entradas conhecidas para prever um valor de resultado ou de resposta. Em muitas aplicações no mundo real, no entanto, os valores de entradas são incertos. A simulação permite considerar a incerteza nas entradas para modelos preditivos e avaliar a probabilidade de vários resultados do modelo na presença de uma incerteza. Por exemplo, você tem um modelo de lucro que inclui o custo de materiais como uma entrada, mas há uma incerteza desse custo devido à volatilidade do mercado. É possível usar simulação para modelar essa incerteza e determinar o efeito que isso causa no lucro.

A simulação no IBM SPSS Statistics utiliza o método de Monte Carlo. Entradas incertas são modeladas com as distribuições de probabilidade (como a distribuição triangular), e valores simulados para essas entradas são gerados ao desenhar a partir dessas distribuições. Entradas cujos valores são conhecidos são mantidas fixas nos valores conhecidos. O modelo preditivo é avaliado utilizando um valor simulado para cada entrada incerta e valores fixos para as entradas conhecidas para calcular a resposta (ou respostas) do modelo. O processo é repetido muitas vezes (normalmente dezenas ou centenas de milhares de vezes), resultando em uma distribuição de valores de resposta que podem ser usados para responder perguntas de uma natureza probabilística. No contexto do IBM SPSS Statistics, cada repetição do processo gera um caso separado (registro) de dados que consiste no conjunto de valores simulados para as entradas incertas, nos valores das entradas fixas e na resposta (ou respostas) previstas do modelo.

Também é possível simular dados na ausência de um modelo preditivo ao especificar distribuições de probabilidade para variáveis que devem ser simuladas. Cada caso de dados gerado consiste no conjunto de valores simulados para as variáveis especificadas.

Para executar uma simulação, é necessário especificar detalhes como o modelo preditivo, as distribuições de probabilidade para as entradas incertas, as correlações entre essas entradas e os valores para quaisquer entradas fixas. Após ter especificado todos os detalhes para uma simulação, será possível executá-la e, opcionalmente, salvar as especificações em um arquivo de **plano de simulação**. É possível



compartilhar o plano de simulação com outros usuários que, em seguida, poderão executar a simulação sem a necessidade de entender os detalhes de como ela foi criada.

Duas interfaces estão disponíveis para trabalhar com simulações. O Construtor de Simulação é uma interface avançada para usuários que estiverem projetando e executando simulações. Ele fornece o conjunto completo de capacidades para projetar uma simulação, salvar as especificações em um arquivo de plano de simulação, especificar a saída e executar a simulação. É possível construir uma simulação com base em um arquivo de modelo do IBM SPSS ou em um conjunto de equações customizadas que você define no Construtor de Simulação. Também é possível carregar um plano de simulação existente no Construtor de Simulação, modificar qualquer uma das configurações e executar a simulação e, opcionalmente, salvar o plano atualizado. Para os usuários que possuem um plano de simulação e sobretudo desejarem executar a simulação, uma interface mais simples está disponível. Com ela é possível modificar as configurações que permitem executar a simulação sob diferentes condições, mas não fornece os recursos completos do Construtor de Simulação para projetar simulações.

## Para designar uma simulação baseada em um arquivo de modelo

1. A partir dos menus, escolha:

**Analisar > Simulação...**

2. Clique em **Selecionar arquivo de modelo SPSS** e clique em **Continuar**.

3. Abra o arquivo de modelo.

O arquivo de modelo é um arquivo XML que contém o modelo PMML criado a partir do IBM SPSS Statistics ou IBM SPSS Modeler. Consulte o tópico [“Guia Modelo”](#) na página 251 para obter mais informações

4. Na guia Simulação (no Construtor de Simulação), especifique as distribuições de probabilidade para entradas e valores simulados para entradas fixas. Se o conjunto de dados ativo contém dados históricos para as entradas simuladas, clique em **Ajustar todos** para determinar automaticamente a distribuição que melhor ajusta os dados a cada entrada, bem como determinar as correlações entre eles. Para cada entrada simulada que não está sendo ajustada aos dados históricos, você deve especificar explicitamente uma distribuição selecionando um tipo de distribuição e inserindo os parâmetros necessários.

5. Clique em **Executar** para executar a simulação. Por padrão, o plano de simulação, especificando os detalhes da simulação, é salvo no local especificado nas configurações Salvar.

As seguintes opções estão disponíveis:

- Modifique a localização do plano de simulação salvo.
- Especifique as correlações conhecidas entre as entradas simuladas.
- Calcule automaticamente uma tabela de contingência de associações entre entradas categóricas e use essas associações quando os dados forem gerados para essas entradas.
- Especifique a análise de sensibilidade para investigar o efeito de variar o valor de uma entrada fixa ou variar um parâmetro de distribuição para uma entrada simulada.
- Especifique opções avançadas, como configurar o número máximo de casos a serem gerados ou solicitar amostragem de cauda.
- Customize a saída.
- Salve os dados simulados em um arquivo de dados.

## Para designar uma simulação baseada em equações customizadas

1. A partir dos menus, escolha:

**Analisar > Simulação...**

2. Clique em **Tipo em equações** e clique em **Continuar**.

3. Clique em **Nova equação** na guia Modelo (no Construtor de Simulação) para definir cada equação em seu modelo preditivo.

4. Clique na guia **Simulação** e especifique distribuições de probabilidade para entradas simuladas e valores para entradas fixas. Se o conjunto de dados ativo contém dados históricos para as entradas simuladas, clique em **Ajustar todos** para determinar automaticamente a distribuição que melhor ajusta os dados a cada entrada, bem como determinar as correlações entre eles. Para cada entrada simulada que não está sendo ajustada aos dados históricos, você deve especificar explicitamente uma distribuição selecionando um tipo de distribuição e inserindo os parâmetros necessários.
5. Clique em **Executar** para executar a simulação. Por padrão, o plano de simulação, especificando os detalhes da simulação, é salvo no local especificado nas configurações Salvar.

As seguintes opções estão disponíveis:

- Modifique a localização do plano de simulação salvo.
- Especifique as correlações conhecidas entre as entradas simuladas.
- Calcule automaticamente uma tabela de contingência de associações entre entradas categóricas e use essas associações quando os dados forem gerados para essas entradas.
- Especifique a análise de sensibilidade para investigar o efeito de variar o valor de uma entrada fixa ou variar um parâmetro de distribuição para uma entrada simulada.
- Especifique opções avançadas, como configurar o número máximo de casos a serem gerados ou solicitar amostragem de cauda.
- Customize a saída.
- Salve os dados simulados em um arquivo de dados.

## Para designar uma simulação sem um modelo preditivo

1. A partir dos menus, escolha:

**Analisar > Simulação...**

2. Clique em **Criar dados simulados** e clique em **Continuar**.
3. Na guia Modelo (no Construtor de simulação), selecione os campos que deseja simular. É possível selecionar campos a partir do conjunto de dados ativo ou definir novos campos clicando em **Novo**.
4. Clique na guia **Simulação** e especifique as distribuições de probabilidade para os campos que devem ser simulados. Se o conjunto de dados ativo contém dados históricos para qualquer um desses campos, clique em **Ajustar todos** para determinar automaticamente a distribuição que melhor ajusta os dados e para determinar as correlações entre os campos. Para campos que não se ajustam aos dados históricos, você deve especificar explicitamente uma distribuição selecionando um tipo de distribuição e inserindo os parâmetros necessários.
5. Clique em **Executar** para executar a simulação. Por padrão, os dados simulados são salvos no novo conjunto de dados especificado nas configurações Salvar. Além disso, o plano de simulação, que especifica os detalhes da simulação, é salvo no local especificado nas configurações Salvar.

As seguintes opções estão disponíveis:

- Modifique a localização dos dados simulados ou o plano de simulação salvo.
- Especifique as correlações conhecidas entre os campos simulados.
- Calcule automaticamente uma tabela de contingência de associações entre campos categóricos e use essas associações quando os dados forem gerados para esses campos.
- Especifique a análise de sensibilidade para investigar o efeito de variar um parâmetro de distribuição para um campo simulado.
- Especifique opções avançadas, como definir o número de casos a serem gerados.

## Para executar uma simulação a partir de um plano de simulação

Duas opções estão disponíveis para executar uma simulação a partir de um plano de simulação. Você pode usar a caixa de diálogo Executar simulação, que foi projetada principalmente para executar a partir de um plano de simulação, ou pode usar o Construtor de Simulação.

Para usar o diálogo Executar Simulação:

1. A partir dos menus, escolha:

**Analisar > Simulação...**

2. Clique em **Abrir um plano de simulação existente**.

3. Certifique-se de que caixa de seleção **Abrir no Construtor de Simulação** não esteja marcada e clique em **Continuar**.

4. Abra o plano de simulação.

5. Clique em **Executar** no diálogo Executar Simulação.

Para executar a simulação a partir do Construtor de Simulação:

1. A partir dos menus, escolha:

**Analisar > Simulação...**

2. Clique em **Abrir um plano de simulação existente**.

3. Marque a caixa de seleção **Abrir no Construtor de Simulação** e clique em **Continuar**.

4. Abra o plano de simulação.

5. Modifique quaisquer configurações que você deseja modificar na guia Simulação.

6. Clique em **Executar** para executar a simulação.

Opcionalmente, é possível fazer o seguinte:

- Configure ou modifique a análise de sensibilidade para investigar o efeito de variar o valor de uma entrada fixa ou variar um parâmetro de distribuição para uma entrada simulada.
- Reajuste distribuições e correlações para entradas simuladas para novos dados.
- Altere a distribuição para uma entrada simulada.
- Customize a saída.
- Salve os dados simulados em um arquivo de dados.

## Construtor de simulação

O construtor de simulação fornece o conjunto completo de recursos para projetar e executar simulações. Permite executar as tarefas gerais a seguir:

- Projetar e executar uma simulação para um modelo do IBM SPSS definido em um arquivo de modelo PMML.
- Projete e execute uma simulação para um modelo preditivo definido por um conjunto de equações customizadas que você especifica.
- Projete e execute uma simulação que gere dados na ausência de um modelo preditivo.
- Execute uma simulação com base em um plano de simulação existente, modificando opcionalmente quaisquer configurações do plano.

## Guia Modelo

Para simulações baseadas em um modelo preditivo, a guia modelo especifica a origem do modelo. Para simulações que não incluem um modelo preditivo, a guia modelo especifica os campos que devem ser simulados.

**Selecione um arquivo de modelo SPSS.** Esta opção especifica que o modelo preditivo é definido em um arquivo de modelo do IBM SPSS. Um arquivo de modelo do IBM SPSS é um arquivo XML ou um archive compactado (arquivo .zip) que contém o modelo PMML criado a partir do IBM SPSS Statistics ou do IBM SPSS Modeler. Os modelos preditivos são criados por procedimentos, como Regressão Linear e Árvores de Decisão, no IBM SPSS Statistics e podem ser exportados em um arquivo de modelo. É possível utilizar um arquivo de modelo diferente clicando em **Procurar** e navegando para o arquivo desejado.

## Modelos PMML suportados pela simulação

- Regressão linear
- Modelo linear automático
- Modelo linear generalizado
- Modelo linear generalizado misto
- Modelo linear geral
- Regressão logística binária
- Regressão logística multinomial
- Regressão Multnomial Ordinal
- Regressão de Cox
- Árvore
- Árvore Impulsionada (C5)
- Discriminante
- Clusterização de duas etapas
- Cluster de k-médias
- Rede neural
- Conjunto de regras (lista de decisão)

### Nota:

- Os modelos PMML que possuem vários campos de destino (variáveis) ou divisões não são suportados para uso na simulação.
- Os valores de entradas de sequência para modelos de regressão logística binária são limitados a 8 bytes no modelo. Se você estiver ajustando essas entradas de sequência ao conjunto de dados ativo, certifique-se de que os valores nos dados não excedam 8 bytes de comprimento. Os valores de dados que excedem 8 bytes são excluídos da distribuição categórica associada para a entrada e são exibidos como não correspondentes na tabela de saída Categorias sem correspondência.

**Tipo nas equações para o modelo.** Esta opção especifica que o modelo preditivo consiste em uma ou mais equações customizadas a serem criadas por você. Crie equações clicando em **Nova Equação**. Isto abre o Editor de Equação. É possível modificar equações existentes, copiá-las para usar como modelos para novas equações, reordená-las e excluí-las.

- O Construtor de Simulação não oferece suporte a sistemas de equações simultâneas ou equações não lineares na variável de destino.
- As equações customizadas são avaliadas na ordem em que são especificadas. Se a equação para um determinado alvo depende de outro destino, então o outro alvo deve ser definido por uma equação anterior.

Por exemplo, dado o conjunto de três equações a seguir, a equação para *profit* depende dos valores de *revenue* e *expenses*, de modo que as equações para *revenue* e *expenses* devem preceder a equação para *profit*.

`revenue = price*volume`

`expenses = fixed + volume*(unit_cost_materials + unit_cost_labor)`

`profit = revenue - expenses`

**Criar dados simulados sem um modelo.** Selecione esta opção para simular dados sem um modelo preditivo. Especifique os campos que devem ser simulados selecionando campos do conjunto de dados ativo ou clicando em **Novo** para definir novos campos.

### Editor de Equação

O Editor de Equação permite criar ou modificar uma equação customizada para seu modelo preditivo.

- A expressão para a equação pode conter campos do conjunto de dados ativo ou novos campos de entrada que você define no Editor de Equação.
  - É possível especificar propriedades da resposta como seu nível de medição, rótulos de valor e se a saída é gerada para a resposta.
  - É possível utilizar respostas de equações definidas anteriormente como entradas para a equação atual, permitindo criar equações acopladas.
  - É possível anexar um comentário descritivo na equação. Os comentários são exibidos junto da equação na guia Modelo.
1. Insira o nome da resposta. Opcionalmente, clique em **Editar** na caixa de texto Resposta para abrir o diálogo Entradas Definidas, permitindo alterar as propriedades padrão da resposta.
  2. Para construir uma expressão, cole componentes no campo Expressão Numérica ou digite diretamente no campo Expressão Numérica.
- É possível construir sua expressão utilizando campos do conjunto de dados ativo ou definir novas entradas clicando no botão **Novo**. Isso abre o diálogo Definir Entradas.
  - É possível colar funções selecionando um grupo da lista de grupos Função e dando um clique duplo na função na lista Funções (ou selecione a função e clique na seta adjacente à lista de grupos Função). Insira quaisquer parâmetros indicados por pontos de interrogação. O grupo de funções rotulado como **Tudo** fornece uma lista de todas as funções disponíveis. Uma breve descrição da função selecionada atualmente é exibida em uma área reservada na caixa de diálogo.
  - Constantes da sequência devem ser colocadas entre aspas.
  - Se os valores contiverem decimais, (.) deve ser usado como o indicador decimal.

*Nota:* A simulação não suporta equações customizadas com respostas de sequência de caracteres.

### *Entradas Definidas*

A caixa de diálogo Entradas Definidas permite definir novas entradas e definir propriedades para destinos.

- Se uma entrada a ser usada em uma equação não existir no conjunto de dados ativo, você deverá defini-la antes que ela possa ser usada na equação.
- Se você estiver simulando dados sem um modelo preditivo, deverá definir todas as entradas simuladas que não existem no conjunto de dados ativo.

**Nome.** Especifique o nome para um destino ou entrada.

**Destino.** É possível especificar um nível de medição para uma resposta. O nível de medição padrão é contínuo. Também é possível especificar se a saída será criada para este destino. Por exemplo, para um conjunto de equações acopladas, você pode estar interessado apenas na saída do destino para a equação final, portanto, suprimiria a saída dos outros destinos.

**Entrada a ser simulada.** Isso especifica que os valores da entrada serão simulados de acordo com uma distribuição de probabilidade especificada (a distribuição de probabilidade é especificada na guia Simulação). O nível de medição determina o conjunto padrão de distribuições que são consideradas ao localizar a distribuição que melhor ajusta os dados para a entrada (ao clicar em **Ajustar** ou **Ajustar Tudo** na guia Simulação). Por exemplo, se o nível de medição for contínuo, então a distribuição normal (apropriada para dados contínuos) seria considerada, mas a distribuição binomial não.

**Nota:** Selecione um nível de medição de String para entradas de sequência de caracteres. As entradas de sequência de caracteres que devem ser simuladas são restritas à distribuição Categórica.

**Entrada de valor fixo.** Isso especifica que o valor da entrada é conhecido e será fixado no valor conhecido. As entradas fixas podem ser numéricas ou sequência de caracteres. Especifique um valor para a entrada fixa. Os valores de sequência de caracteres não devem ser colocados entre aspas.

**Rótulos de valor.** É possível especificar rótulos de valor para respostas, entradas simuladas e entradas fixas. Os rótulos de valor são usados em gráficos e tabelas de saída.

## Guia Simulação

A guia Simulação especifica todas as propriedades da simulação, exceto o modelo preditivo. É possível executar as seguintes tarefas gerais na guia Simulação:

- Especifique distribuições de probabilidade para entradas simuladas e valores para entradas fixas.
- Especificar as correlações entre as entradas simuladas. Para entradas categóricas, você pode especificar que as associações que existem entre essas entradas no conjunto de dados ativo sejam usadas quando os dados forem gerados para essas entradas.
- Especifique opções avançadas, como amostragem de cauda e critérios para ajustar as distribuições aos dados históricos.
- Customize a saída.
- Especifique onde salvar o plano de simulação e, opcionalmente, salve os dados simulados.

### Campos Simulados

Para executar uma simulação, cada campo de entrada deve ser especificado como fixo ou simulado. As entradas simuladas são aquelas cujos valores são incertos e serão geradas a partir de uma distribuição de probabilidade especificada. Quando os dados históricos estão disponíveis para as entradas a serem simuladas, as distribuições que melhor se ajustarem aos dados podem ser determinadas automaticamente, junto de quaisquer correlações entre essas entradas. Também é possível especificar manualmente distribuições ou correlações se os dados históricos não estiverem disponíveis ou se você precisar de distribuições ou correlações específicas.

As entradas fixas são aquelas cujos valores são conhecidos e permanecem constantes para cada caso gerado na simulação. Por exemplo, você possui um modelo de regressão linear para vendas como uma função de um número de entradas que incluem preço, e você deseja manter o preço fixo ao preço de mercado atual. Você, então, especifica o preço como uma entrada fixa.

Para simulações com base em um modelo preditivo, cada um preditor no modelo é um campo de entrada para a simulação. Para simulações que não incluem um modelo preditivo, os campos que são especificados na guia Modelo são as entradas para a simulação.

### Ajustando automaticamente as distribuições e calculando as correlações para as entradas simuladas.

Se o conjunto de dados ativo contiver dados históricos para as entradas que deseja simular, então será possível localizar automaticamente as distribuições que melhor ajustam os dados para essas entradas, além de determinar quaisquer correlações entre elas. As etapas são as seguintes:

1. Verifique se cada uma das entradas que deseja simular é correspondida com o campo correto no conjunto de dados ativo. As entradas são listadas na coluna de entrada e a coluna Ajustar a exibe o campo correspondente no conjunto de dados ativo. É possível corresponder uma entrada a um campo diferente no conjunto de dados ativo ao selecionar um item diferente na lista suspensa Ajustar a.

Um valor de *-None-* na coluna Ajustar a indica que a entrada não pôde ser correspondida automaticamente a um campo no conjunto de dados ativo. Por padrão, as entradas são correspondidas aos campos de conjunto de dados por nome, nível de medição e tipo (numérico ou sequência de caracteres). Se o conjunto de dados ativo não contiver dados históricos para a entrada, então especifique manualmente a distribuição para a entrada ou especifique a entrada como uma entrada fixa, conforme descrito abaixo.

2. Clique em **Ajustar Tudo**.

A distribuição de ajuste mais próxima e seus parâmetros associados são exibidos na coluna Distribuição junto de um gráfico da distribuição sobreposta em um histograma (ou gráfico de barras) dos dados históricos. As correlações entre entradas simuladas são exibidas nas configurações de Correlações. É possível examinar os resultados do ajuste e customizar o ajuste de distribuição automático para uma entrada específica ao selecionar a linha para a entrada e clicar em **Detalhes do Ajuste**. Consulte o tópico [“Detalhes do Ajuste”](#) na página 256 para obter mais informações.

É possível executar ajuste de distribuição automático para uma entrada específica ao selecionar a linha para a entrada e clicar em **Ajustar**. Correlações para todas as entradas simuladas que correspondem aos campos no conjunto de dados ativo também são calculadas automaticamente.

**Nota:**

- Casos com valores ausentes para qualquer entrada simulada são excluídos do ajuste de distribuição, cálculo de correlações e cálculo da tabela de contingência opcional (para entradas com distribuição Categórica). Opcionalmente, você pode especificar se os valores ausentes de usuário de entradas com uma distribuição Categórica são tratados como válidos. Por padrão, eles são tratados como ausentes. Para obter mais informações, consulte o tópico “Opções Avançadas ” na página 258.
- Para entradas contínuas e ordinais, se um ajuste aceitável não puder ser encontrado para qualquer uma das distribuições testadas, a distribuição Empírica é sugerida como o ajuste mais próximo. Para entradas contínuas, a distribuição empírica é a função de distribuição cumulativa dos dados históricos. Para entradas ordinais, a distribuição Empírica é a distribuição categórica dos dados históricos.

**Especificando distribuições manualmente.** É possível especificar manualmente a distribuição de probabilidade para qualquer entrada simulada ao selecionar a distribuição na lista suspensa **Tipo** e inserir os parâmetros de distribuição na grade Parâmetros. Após ter inserido os parâmetros para uma distribuição, um gráfico de amostra da distribuição, com base nos parâmetros especificados, será exibido ao lado da grade Parâmetros. A seguir há algumas notas sobre distribuições específicas:

- **Categórico.** A distribuição categórica descreve um campo de entrada que possui um número fixo de valores, referidos como categorias. Cada categoria possui uma probabilidade associada, de forma que a soma das probabilidades de todas as categorias seja igual a um. Para inserir uma categoria, clique na coluna à esquerda na grade Parâmetros e especifique o valor de categoria. Insira a probabilidade associada à categoria na coluna à direita.

**Nota:** Entradas categóricas a partir de um modelo PMML possuem categorias que são determinadas a partir do modelo e não podem ser modificadas.

- **Binomial Negativa - Falhas.** Descreve a distribuição do número de falhas em uma sequência de avaliações antes que um número especificado de sucessos seja observado. O parâmetro *thresh* é o número especificado de sucessos e o parâmetro *prob* é a probabilidade de sucesso em qualquer avaliação determinada.
- **Binomial Negativa - Avaliações.** Descreve a distribuição do número de avaliações necessárias antes que um número especificado de sucessos seja observado. O parâmetro *thresh* é o número especificado de sucessos e o parâmetro *prob* é a probabilidade de sucesso em qualquer avaliação determinada.
- **Range.** Esta distribuição consiste em um conjunto de intervalos com uma probabilidade designada a cada intervalo, de tal forma que a soma das probabilidades sobre todos os intervalos é igual a 1. Os valores dentro de um determinado intervalo são obtidos a partir de uma distribuição uniforme definida nesse intervalo. Os intervalos são especificados ao inserir um valor mínimo, um valor máximo e uma probabilidade associada.

Por exemplo, você acredita que o custo de uma matéria-prima tem 40% de chance de cair na faixa de \$10 - \$15 por unidade e de 60% de chance de cair na faixa de \$15 - \$20 por unidade. Você modela o custo com uma distribuição de Intervalo consistindo em dois intervalos [10 – 15] e [15 – 20], configurando a probabilidade associada ao primeiro intervalo para 0,4 e a probabilidade associada ao segundo intervalo para 0,6. Os intervalos não precisam ser contínuos e podem até mesmo ser sobrepostos. Por exemplo, você poderia ter especificado os intervalos \$10 - \$15 e \$20 - \$25 ou \$10 - \$15 e \$13 - \$16.

- **Weibull.** O parâmetro *c* é um parâmetro de localização opcional, que especifica onde a origem da distribuição está localizada.

Os parâmetros para as seguintes distribuições têm o mesmo significado que nas funções de variáveis aleatórias associadas disponíveis na caixa de diálogo Calcular Variável: Bernoulli, beta, binomial, exponencial, gama, lognormal, binomial negativa (tentativas e falhas), normal, Poisson e uniforme.

**Especificando entradas fixas.** Especifique uma entrada fixa ao selecionar Fixo a partir da lista suspensa **Tipo** na coluna Distribuição e inserir o valor fixo. O valor pode ser numérico ou sequência de caracteres, dependendo se a entrada é numérica ou uma sequência de caracteres. Os valores de sequência não devem ser colocados entre aspas.

**Especificando limites em valores simulados.** A maioria das distribuições suporta a especificação de limites superior e inferior em valores simulados. É possível especificar um limite inferior ao inserir um valor na caixa de texto **Mín.**, e especificar um limite superior ao inserir um valor na caixa de texto **Máx.**.






**Bloqueando entradas.** Bloquear uma entrada, ao marcar a caixa de seleção na coluna com o ícone de bloqueio, exclui a entrada do ajuste distribuição automático. Isso é mais útil ao especificar manualmente uma distribuição ou um valor fixo e desejar assegurar que essa entrada não seja afetada pelo ajuste de distribuição automático. O bloqueio também será útil se pretender compartilhar seu plano de simulação com usuários que irão executá-lo no diálogo Executar Simulação e desejar impedir quaisquer mudanças em determinadas entradas. Neste sentido, as especificações para entradas bloqueadas não podem ser modificadas no diálogo Executar Simulação.

**Análise de sensibilidade.** A análise de sensibilidade permite investigar o efeito das mudanças sistemáticas em uma entrada fixa ou em um parâmetro de distribuição para uma entrada simulada ao gerar um conjunto independente de casos simulados – efetivamente, uma simulação separada - para cada valor especificado. Para especificar análise de sensibilidade, selecione uma entrada fixa ou simulada e clique em **Análise de Sensibilidade**. A análise de sensibilidade é limitada a uma única entrada fixa ou a um único parâmetro de distribuição para uma entrada simulada. Consulte o tópico “Análise de Sensibilidade” na página 257 para obter informações adicionais.

Ícone de status de Ajuste

Os ícones na coluna Ajustar a indicam o status de ajuste para cada campo de entrada.

Tabela 5. Ícones de status

Ícone	Descrição
	Nenhuma distribuição foi especificada para a entrada, e a entrada não foi especificada como fixa. Para executar a simulação, deve-se especificar uma distribuição para esta entrada ou defini-la para ser fixa e especificar o valor fixo.
	A entrada foi ajustada anteriormente para um campo que não existe no conjunto de dados ativo. Nenhuma ação é necessária, a menos que você deseje reajustar a distribuição para a entrada no conjunto de dados ativo.
	A distribuição de ajuste mais próxima foi substituída por uma distribuição alternativa no diálogo Detalhes do Ajuste.
	A entrada é configurada como a distribuição de ajuste mais próxima.
	A distribuição foi especificada manualmente ou as iterações da análise de sensibilidade foram especificadas para essa entrada.

#### Detalhes do Ajuste

O diálogo Detalhes do Ajuste exibe os resultados do ajuste de distribuição automático para uma entrada específica. As distribuições são ordenadas pela Qualidade do ajuste, com a distribuição de ajuste mais próxima listada primeiro. É possível substituir a distribuição de ajuste mais próxima ao selecionar o botão de opções para a distribuição desejada na coluna Usar. Selecionar um botão de opções na coluna Usar também exibe um gráfico da distribuição sobreposto em um histograma (ou um gráfico de barras) dos dados históricos para essa entrada.

**Estatística de ajuste.** Por padrão e para campos contínuos, o teste de Anderson-Darling é utilizado para determinar a Qualidade do ajuste. Como alternativa e somente para campos contínuos, é possível especificar o teste de Kolmogorov-Smirnoff para Qualidade do ajuste ao selecionar essa opção nas configurações de Opções Avançadas. Para entradas contínuas, os resultados dos dois testes são mostrados na coluna Estatísticas de Ajuste (A para Anderson-Darling e K para Kolmogorov-Smirnoff), com o teste escolhido utilizado para ordenar as distribuições. Para entradas ordinais e nominais, o teste qui-quadrado é utilizado. Os valores p associados aos testes também são mostrados.



**Parâmetros.** Os parâmetros de distribuição associados a cada distribuição ajustada são exibidos na coluna Parâmetros. Os parâmetros para as seguintes distribuições possuem o mesmo significado que as funções de variável aleatórias associadas disponíveis na caixa de diálogo Calcular Variável: Bernoulli, beta, binomial, exponencial, gama, lognormal, binomial negativa (avaliações e falhas), normal, Poisson e uniforme. Para a distribuição categórica, os nomes dos parâmetros são as categorias e os valores dos parâmetros são as probabilidades associadas.

**Reajustando com um conjunto de distribuição customizado.** Por padrão, o nível de medição da entrada é utilizado para determinar o conjunto de distribuições consideradas para ajuste de distribuição automático. Por exemplo, distribuições contínuas como lognormal e gama são consideradas ao ajustar uma entrada contínua, mas distribuições discretas, como Poisson e binomial, não são. É possível escolher um subconjunto das distribuições padrão ao selecionar as distribuições na coluna Reajustar. Também é possível substituir o conjunto padrão das distribuições ao selecionando um nível de medição diferente da lista suspensa **Tratar como (Medida)** e selecionar as distribuições na coluna Reajustar. Clique em **Executar Reajuste** para reajustar com o conjunto de distribuição customizado.

**Nota:**

- Casos com valores ausentes para qualquer entrada simulada são excluídos do ajuste de distribuição, cálculo de correlações e cálculo da tabela de contingência opcional (para entradas com distribuição Categórica). Opcionalmente, você pode especificar se os valores ausentes de usuário de entradas com uma distribuição Categórica são tratados como válidos. Por padrão, eles são tratados como ausentes. Para obter mais informações, consulte o tópico “Opções Avançadas ” na página 258.
- Para entradas contínuas e ordinais, se um ajuste aceitável não puder ser encontrado para qualquer uma das distribuições testadas, a distribuição Empírica é sugerida como o ajuste mais próximo. Para entradas contínuas, a distribuição empírica é a função de distribuição cumulativa dos dados históricos. Para entradas ordinais, a distribuição Empírica é a distribuição categórica dos dados históricos.

*Análise de Sensibilidade*

A análise de sensibilidade permite investigar o efeito de variar uma entrada fixa ou um parâmetro de distribuição para uma entrada simulada sobre um conjunto de valores especificado. Um conjunto independente de casos simulados - efetivamente, uma simulação separada - é gerado para cada valor especificado, permitindo que você investigue o efeito da variação da entrada. Cada conjunto de casos simulados é referido como uma **iteração**.

**Iterar.** Esta escolha permite especificar o conjunto de valores sobre os quais a entrada será variada.

- Se você estiver variando o valor de um parâmetro de distribuição, selecione o parâmetro na lista suspensa. Insira o conjunto de valores na grade Valor do parâmetro por iteração. Clicar em **Continuar** incluirá os valores especificados na grade Parâmetros da entrada associada, com um índice especificando o número de iteração do valor.
- Para as distribuições Categórica e Intervalo, as probabilidades das categorias ou intervalos, respectivamente, podem ser variadas, mas os valores das categorias e os pontos finais dos intervalos não podem ser variados. Selecione uma categoria ou intervalo na lista suspensa e especifique o conjunto de probabilidades no valor do parâmetro por grade de iteração. As probabilidades para as outras categorias ou intervalos serão ajustadas de modo automático e apropriado.

**Nenhuma iteração.** Utilize essa opção para cancelar as iterações para uma entrada. Clicar em **Continuar** removerá as iterações.

**Correlações**

Os campos de entrada a serem simulados geralmente são correlacionados, por exemplo, a altura e o peso. As correlações entre as entradas que serão simuladas devem ser contabilizadas para garantir que os valores simulados preservem essas correlações.

**Recalcular correlações ao ajustar.** Essa opção especifica que as correlações entre entradas simuladas são calculadas automaticamente ao ajustar as distribuições para o conjunto de dados ativo por meio das ações **Ajustar Tudo** ou **Ajustar** nas configurações de Campos Simulados.

**Não recalculer correlações ao ajustar.** Selecione esta opção se desejar especificar correlações manualmente e evitar que sejam substituídas ao ajustar automaticamente as distribuições ao conjunto de dados ativo. Os valores inseridos na grade Correlações devem estar entre -1 e 1. Um valor de 0 especifica que não há correlação entre o par de entradas associado.

**Reconfigurar.** Isso redefine todas as correlações para 0.

**Use a tabela de contingência multiponto ajustada para entradas com distribuição categórica.**

Para entradas com uma distribuição categórica, é possível calcular automaticamente uma tabela de contingência multidirecional a partir do conjunto de dados ativo que descreve as associações entre essas entradas. A tabela de contingência é então usada quando os dados são gerados para essas entradas. Se você optar por salvar o plano de simulação, a tabela de contingência será salva no arquivo de plano e será usada ao executar o plano.

- **Calcular a tabela de contingência a partir do conjunto de dados ativo.** Se você estiver trabalhando com um plano de simulação existente que contém uma tabela de contingência, poderá recalculá-la a partir do conjunto de dados ativo. Esta ação substitui a tabela de contingência do arquivo de plano carregado.
- **Utilizar tabela de contingência a partir do plano de simulação carregado.** Por padrão, quando você carrega um plano de simulação que contém uma tabela de contingência, a tabela do plano é usada. É possível recalculá-la a partir do conjunto de dados ativo ao selecionar **Calcular tabela de contingência a partir do conjunto de dados ativo**.

### **Opções Avançadas**

**Número máximo de casos.** Isso especifica o número máximo de casos de dados simulados e valores de destino associados a serem gerados. Quando a análise de sensibilidade é especificada, esse é o número máximo de casos para cada iteração.

**Resposta para critérios de parada.** Se seu modelo preditivo contiver mais de um destino, você poderá selecionar o destino ao qual os critérios de parada serão aplicados.

**Critérios de parada.** Essas escolhas especificam critérios para parar a simulação, potencialmente antes que o número máximo de casos permitidos seja gerado.

- **Continuar até que o máximo seja atingido.** Isso especifica que os casos simulados serão gerados até que o número máximo de casos seja atingido.
- **Parar quando as caudas tiverem sido amostradas.** Use esta opção quando desejar garantir que uma das caudas de uma distribuição de destino especificada tenha sido amostrada adequadamente. Os casos simulados serão gerados até que a amostragem de cauda especificada seja concluída ou o número máximo de casos seja atingido. Se o seu modelo preditivo contiver diversas respostas, então selecione a resposta à qual estes critérios serão aplicados, na lista suspensa **Resposta para critérios de parada**.

**Tipo.** Você pode definir o limite da região da cauda especificando um valor do destino como 10.000.000 ou um percentil como o 99º percentil. Se escolher Valor na lista suspensa **Tipo**, insira o valor do limite na caixa de texto Valor e utilize a lista suspensa **Lado** para especificar se este é o limite da região da cauda Esquerda ou da região da cauda Direita. Se escolher Percentil na lista suspensa **Tipo**, insira um valor na caixa de texto Percentil.

**Frequência.** Especifique o número de valores do destino que devem estar na região da cauda para garantir que a cauda foi amostrada adequadamente. A geração de casos será interrompida quando esse número for atingido.

- **Parar quando o intervalo de confiança da média estiver dentro do limite especificado.** Use esta opção quando quiser garantir que a média de um determinado alvo seja conhecida com um grau de precisão especificado. Casos simulados serão gerados até que o grau de precisão especificado seja alcançado ou o número máximo de casos seja alcançado. Para usar essa opção, você especifica um nível de confiança e um limite. Casos simulados serão gerados até que o intervalo de confiança associado ao nível especificado esteja dentro do limite. Por exemplo, é possível usar esta opção para especificar que os casos sejam gerados até que o intervalo de confiança da média no nível de confiança de 95% esteja dentro de 5% do valor médio. Se o seu modelo preditivo contiver diversas respostas,

então selecione a resposta à qual estes critérios serão aplicados, na lista suspensa **Resposta para critérios de parada**.

**Tipo de Limite.** É possível especificar o limite como um valor numérico ou como uma porcentagem da média. Se escolher Valor na lista suspensa **Tipo de Limite**, insira um limite na caixa de texto Limite como Valor. Se escolher Porcentagem na lista suspensa **Tipo do Limite**, insira um valor na caixa de texto Limite como Porcentagem.

**Número de casos para amostra.** Isso especifica o número de casos a serem usados ao ajustar automaticamente as distribuições para entradas simuladas no conjunto de dados ativo. Se seu conjunto de dados for muito grande, você pode considerar limitar o número de casos usados para ajuste de distribuição. Se selecionar **Limitar para N casos**, os primeiros N casos serão utilizados.

**Crítérios de Qualidade do ajuste (Contínuos).** Para entradas contínuas, é possível usar o teste de Anderson-Darling ou o teste Kolmogorov-Smirnoff de qualidade de ajuste para classificar distribuições ao ajustar distribuições para entradas simuladas ao conjunto de dados ativo. O teste de Anderson-Darling é selecionado por padrão e é especialmente recomendado quando você deseja garantir o melhor ajuste possível nas regiões da cauda.

**Distribuição Empírica.** Para entradas contínuas, a distribuição empírica é a função de distribuição cumulativa dos dados históricos. É possível especificar o número de compartimentos usados para calcular a distribuição empírica para entradas contínuas. O padrão é 100 e o máximo é 1000.

**Replicar resultados.** Configurar uma semente aleatória permite replicar sua simulação. Especifique um número inteiro ou clique em **Gerar**, que criará um pseudonúmero inteiro aleatório entre 1 e 2147483647, inclusive. O padrão é 629111597.

**Nota:** Para uma determinada semente aleatória, os resultados são replicados, a menos que o número de encadeamentos seja alterado. Em um computador específico, o número de encadeamentos é constante a menos que o mude executando a sintaxe de comando SET THREADS. O número de encadeamentos pode mudar se você executar a simulação em um computador diferente, pois um algoritmo interno é usado para determinar o número de encadeamentos para cada computador.

**Valores omissos do usuário para entradas com uma Distribuição Categórica.** Esses controles especificam se os valores omissos do usuário de entradas com uma distribuição Categórica são tratados como válidos. Os valores omissos do sistema e os valores omissos do usuário para todos os outros tipos de entradas são sempre tratados como inválidos. Todas as entradas devem ter valores válidos para que um caso seja incluído no ajuste de distribuição, cálculo de correlações e cálculo da tabela de contingência opcional.

## ***Funções de Densidade***

Essas configurações permitem que você personalize a saída para funções de densidade de probabilidade e funções de distribuição cumulativa para alvos contínuos, bem como gráficos de barras de valores previstos para alvos categóricos.

**Função de Densidade de Probabilidade (PDF).** A função de densidade de probabilidade exibe a distribuição dos valores de destino. Para destinos contínuos, permite determinar a probabilidade de que o alvo esteja dentro de uma determinada região. Para destinos categóricos (destinos com nível de medição nominal ou ordinal), é gerado um gráfico de barras que exibe a porcentagem de casos que se enquadram em cada categoria do destino. Opções adicionais para destinos categóricos de modelos PMML estão disponíveis com a configuração Valores de categoria para relatório descrita abaixo.

Para modelos de cluster de duas etapas e modelos de cluster K-médias, um gráfico de barras de associação de cluster é produzido.

**Função de Distribuição Cumulativa (CDF).** A função de distribuição acumulativa exibe a probabilidade de que o valor da resposta seja menor ou igual a um valor especificado. Está disponível apenas para destinos contínuos.

**Posições da régua de controle.** É possível especificar as posições iniciais das linhas de referência móveis em gráficos PDF e CDF. Os valores especificados para as linhas inferior e superior referem-se a posições ao longo do eixo horizontal, não a percentis. É possível remover a linha inferior ao selecionar **-Infinito** ou

remover a linha superior ao selecionar **Infinito**. Por padrão, as linhas são posicionadas nos percentis 5 e 95. Quando várias funções de distribuição são exibidas em um único gráfico (devido a vários destinos ou resultados de iterações de análise de sensibilidade), o padrão se refere à distribuição para a primeira iteração ou primeiro destino.

**Linhas de Referência (Contínuas).** É possível solicitar que várias linhas de referência verticais sejam adicionadas às funções de densidade de probabilidade e funções de distribuição cumulativa para alvos contínuos.

- **Sigmas.** É possível incluir linhas de referência em mais e menos um número especificado de desvios padrão da média de uma resposta.
- **Percentis.** É possível incluir linhas de referência em um ou dois valores de percentil da distribuição de uma resposta ao inserir valores nas caixas de texto Inferior e Superior. Por exemplo, um valor de 95 na caixa de texto Superior representa o 95º percentil, que é o valor abaixo do qual 95% das observações caem. Da mesma forma, um valor de 5 na caixa de texto Inferior representa o 5º percentil, que é o valor abaixo do qual 5% das observações caem.
- **Linhas de referência customizadas.** É possível incluir linhas de referência em valores especificados do destino.

**Nota:** Quando várias funções de distribuição são exibidas em um único gráfico (devido a vários destinos ou resultados de iterações de análise de sensibilidade), as linhas de referência são aplicadas apenas à distribuição para a primeira iteração ou primeiro destino. É possível incluir linhas de referência a outras distribuições na caixa de diálogo Opções de gráfico, que é acessada a partir do gráfico PDF ou CDF.

**Sobrepôr resultados a partir de variáveis resposta contínuas separadas.** No caso de vários destinos contínuos, isso especifica se as funções de distribuição para todos esses destinos são exibidas em um único gráfico, com um gráfico para funções de densidade de probabilidade e outro para funções de distribuição cumulativa. Quando esta opção não estiver selecionada, os resultados de cada destino serão exibidos em um gráfico separado.

**Valores de categoria para relatório.** Para modelos PMML com destinos categóricos, o resultado do modelo é um conjunto de probabilidades previstas, uma para cada categoria, de que o valor alvo caia em cada categoria. A categoria com a maior probabilidade é considerada a categoria predita e utilizada na geração do gráfico de barras descrito para a configuração **Função de Densidade de Probabilidade** acima. Selecionar **Categoria predita** gera o gráfico de barras. Selecionar **Probabilidades previstas** gera histogramas da distribuição de probabilidades previstas para cada uma das categorias da resposta.

**Agrupamento para análise de sensibilidade.** As simulações que incluem análise de sensibilidade geram um conjunto independente de valores alvo previstos para cada iteração definida pela análise (uma iteração para cada valor da entrada que está sendo variada). Quando as iterações estão presentes, o gráfico de barras da categoria prevista para um destino categórico é exibido como um gráfico de barras agrupado que inclui os resultados de todas as iterações. É possível optar por agrupar categorias ou agrupar iterações.

## Saída

**Gráficos Tornados.** Os gráficos tornado são gráficos de barras que exibem relacionamentos entre destinos e entradas simuladas usando uma variedade de métricas.

- **Correlação da resposta com a entrada.** Esta opção cria um gráfico tornado dos coeficientes de correlação entre um determinado alvo e cada uma de suas entradas simuladas. Este tipo de gráfico tornado não suporta alvos com nível de medição nominal ou ordinal ou entradas simuladas com distribuição categórica.
- **Contribuição para variância.** Essa opção cria um gráfico tornado que exhibe a contribuição para a variação de um destino de cada uma de suas entradas simuladas, permitindo avaliar o grau em que cada entrada contribui para a incerteza geral no destino. Este tipo de gráfico tornado não suporta alvos com níveis de medição ordinais ou nominais ou entradas simuladas com qualquer uma das seguintes distribuições: categórica, Bernoulli, binomial, Poisson ou binomial negativa.
- **Sensibilidade de resposta para alterar.** Esta opção cria um gráfico tornado que exhibe o efeito no alvo de modular cada entrada simulada por mais ou menos um número especificado de desvios padrão da

distribuição associada à entrada. Este tipo de gráfico tornado não suporta alvos com níveis de medição ordinais ou nominais ou entradas simuladas com qualquer uma das seguintes distribuições: categórica, Bernoulli, binomial, Poisson ou binomial negativa.

**Box plots de distribuições de resposta.** Os gráficos de caixa estão disponíveis para alvos contínuos. Selecione **Sobrepôr resultados a partir de respostas separadas** se o seu modelo preditivo possuir diversas variáveis resposta contínuas e você desejar exibir box plots para todas as respostas em um único gráfico.

**Gráficos de dispersão de respostas versus entradas.** Gráficos de dispersão de destinos versus entradas simuladas estão disponíveis para destinos contínuos e categóricos e incluem dispersões do destino com entradas contínuas e categóricas. Dispersões envolvendo um alvo categórico ou uma entrada categórica são exibidas como um mapa de calor.

**Criar uma tabela de valores de percentil.** Para alvos contínuos, é possível obter uma tabela de percentis especificados das distribuições de destino. Os quartis (os percentis 25, 50 e 75) dividem as observações em quatro grupos de igual tamanho. Se desejar um número igual de grupos diferente de quatro, selecione **Intervalos** e especifique o número. Selecione **Percentis Customizados** para especificar percentis individuais - por exemplo, o 99º percentil.

**Estatísticas descritivas de distribuições de resposta.** Esta opção cria tabelas de estatísticas descritivas para metas contínuas e categóricas, bem como para entradas contínuas. Para alvos contínuos, a tabela inclui a média, desvio padrão, mediana, mínimo e máximo, intervalo de confiança da média no nível especificado e os percentis 5 e 95 da distribuição alvo. Para destinos categóricos, a tabela inclui a porcentagem de casos que se enquadram em cada categoria do alvo. Para alvos categóricos de modelos PMML, a tabela também inclui a probabilidade média de cada categoria do alvo. Para entradas contínuas, a tabela inclui a média, desvio padrão, mínimo e máximo.

**Correlações e tabela de contingência para entradas.** Esta opção exibe uma tabela de coeficientes de correlação entre as entradas simuladas. Quando entradas com distribuições categóricas são geradas a partir de uma tabela de contingência, a tabela de contingência dos dados gerados para essas entradas também é exibida.

**Entradas simuladas para inclusão na saída.** Por padrão, todas as entradas simuladas são incluídas na saída. Você pode excluir entradas simuladas selecionadas da saída. Isso os excluirá dos gráficos de tornado, gráficos de dispersão e resultado tabular.

**Limitar intervalos para variáveis resposta contínuas.** É possível especificar o intervalo de valores válidos para um ou mais destinos contínuos. Os valores fora do intervalo especificado são excluídos de todas as saídas e análises associadas aos destinos. Para configurar um limite inferior, selecione **Inferior** na coluna Limite e insira um valor na coluna Mínimo. Para configurar um limite superior, selecione **Superior** na coluna Limite e insira um valor na coluna Máximo. Para configurar um limite inferior e um limite superior, selecione **Ambos** na coluna de Limite e insira os valores nas colunas Mínimo e Máximo.

**Exibir Formatos.** É possível configurar o formato usado ao exibir valores de destinos e entradas (entradas fixas e entradas simuladas).

## Salvar

**Salvar o plano para esta simulação.** É possível salvar as especificações atuais para a sua simulação em um arquivo de plano de simulação. Os arquivos de plano de simulação possuem a extensão *.splan*. É possível reabrir o plano no Construtor de Simulação e, opcionalmente, fazer modificações e executar a simulação. É possível compartilhar o plano de simulação com outros usuários, que podem executá-lo na caixa de diálogo Executar simulação. Os planos de simulação incluem todas as especificações, exceto as seguintes: configurações para funções de densidade; Configurações de saída para gráficos e tabelas; Configurações de opções avançadas para ajuste, distribuição empírica e semente aleatória.

**Salvar os dados simulados como um novo arquivo de dados.** É possível salvar as entradas simuladas, entradas fixas e valores de resposta preditos em um arquivo de dados do SPSS Statistics, em um novo conjunto de dados na sessão atual ou em um arquivo Excel. Cada caso (ou linha) do arquivo de dados consiste nos valores previstos dos alvos junto com as entradas simuladas e entradas fixas que geram os

valores alvo. Quando a análise de sensibilidade é especificada, cada iteração dá origem a um conjunto contíguo de casos que são rotulados com o número da iteração.

## Executar diálogo de simulação

A caixa de diálogo Executar simulação foi projetada para usuários que têm um plano de simulação e desejam principalmente executar a simulação. Ele também fornece os recursos necessários para executar a simulação sob diferentes condições. Permite executar as tarefas gerais a seguir:

- Configure ou modifique a análise de sensibilidade para investigar o efeito de variar o valor de uma entrada fixa ou variar um parâmetro de distribuição para uma entrada simulada.
- Reajuste as distribuições de probabilidade para entradas incertas (e correlações entre essas entradas) para novos dados.
- Modifique a distribuição para uma entrada simulada.
- Customize a saída.
- Execute a simulação.

## Guia de simulação

A guia Simulação permite especificar a análise de sensibilidade, reajustar distribuições de probabilidade para entradas simuladas e correlações entre entradas simuladas para novos dados e modificar a distribuição de probabilidade associada a uma entrada simulada.

A grade de entradas simuladas contém uma entrada para cada campo de entrada definido no plano de simulação. Cada entrada exibe o nome da entrada e o tipo de distribuição de probabilidade associado à entrada, juntamente com um gráfico de amostra da curva de distribuição associada. Cada entrada também tem um ícone de status associado (um círculo colorido com uma marca de seleção) que é útil quando você está reajustando distribuições para novos dados. Além disso, as entradas podem incluir um ícone de cadeado que indica que a entrada está bloqueada e não pode ser modificada ou reajustada para novos dados na caixa de diálogo Executar Simulação. Para modificar uma entrada bloqueada, será necessário abrir o plano de simulação no Construtor de Simulação.

Cada entrada será simulada ou fixa. As entradas simuladas são aquelas cujos valores são incertos e serão gerados a partir de uma distribuição de probabilidade especificada. As entradas fixas são aquelas cujos valores são conhecidos e permanecem constantes para cada caso gerado na simulação. Para trabalhar com uma entrada específica, selecione a entrada para a entrada na grade de entradas simuladas.

Especificando a análise de sensibilidade

A análise de sensibilidade permite investigar o efeito das mudanças sistemáticas em uma entrada fixa ou em um parâmetro de distribuição para uma entrada simulada ao gerar um conjunto independente de casos simulados – efetivamente, uma simulação separada - para cada valor especificado. Para especificar análise de sensibilidade, selecione uma entrada fixa ou simulada e clique em **Análise de Sensibilidade**. A análise de sensibilidade é limitada a uma única entrada fixa ou a um único parâmetro de distribuição para uma entrada simulada. Consulte o tópico [“Análise de Sensibilidade” na página 257](#) para obter informações adicionais.

Reajustando distribuições para novos dados

Para reajustar automaticamente as distribuições de probabilidade para entradas simuladas (e correlações entre entradas simuladas) para dados no conjunto de dados ativo:

1. Verifique se cada uma das entradas do modelo corresponde ao campo correto no conjunto de dados ativo. Cada entrada simulada é ajustada para o campo no conjunto de dados ativo especificado na lista suspensa **Campo** associada a essa entrada. É possível identificar facilmente entradas que não têm correspondência procurando entradas com um ícone de status que inclui uma marca de seleção com um ponto de interrogação, conforme mostrado abaixo.



2. Modifique qualquer correspondência de campo necessária ao selecionar **Ajustar para um campo no conjunto de dados** e selecionar o campo na lista.
3. Clique em **Ajustar Tudo**.

Para cada entrada que foi ajustada, a distribuição que melhor se ajusta aos dados é exibida junto com um gráfico da distribuição sobreposta em um histograma (ou gráfico de barras) dos dados históricos. Se um ajuste aceitável não for localizado, a distribuição empírica é usada. Para entradas que se ajustem à distribuição Empírica, você verá apenas um histograma dos dados históricos porque a distribuição Empírica é de fato representada por esse histograma.

*Nota:* Para obter uma lista completa de ícones de status, consulte o tópico [“Campos Simulados”](#) na página 254.

Modificando distribuições de probabilidade

É possível modificar a distribuição de probabilidade para uma entrada simulada e, opcionalmente, alterar uma entrada simulada para uma entrada fixa ou vice-versa.

1. Selecione a entrada e selecione **Configurar manualmente a distribuição**.
2. Selecione o tipo de distribuição e especifique os parâmetros de distribuição. Para alterar uma entrada simulada para uma entrada fixa, selecione Fixo na lista suspensa **Tipo**.

Depois de inserir os parâmetros de uma distribuição, o gráfico de amostra da distribuição (exibido na entrada da entrada) será atualizado para refletir suas alterações. Para obter mais informações sobre como especificar manualmente as distribuições de probabilidade, consulte o tópico [“Campos Simulados”](#) na página 254.

**Incluir valores omissos de usuário de entradas categóricas ao ajustar.** Isso especifica se os valores omissos do usuário de entradas com uma distribuição Categórica são tratados como válidos quando você está reajustando os dados no conjunto de dados ativo. Os valores omissos do sistema e os valores ausentes do usuário para todos os outros tipos de entradas são sempre tratados como inválidos. Todas as entradas devem ter valores válidos para que um caso seja incluído no ajuste de distribuição e cálculo de correlações.

## Guia de saída

A guia Saída permite personalizar a saída gerada pela simulação.

**Funções de Densidade.** As funções de densidade são o principal meio de sondar o conjunto de resultados de sua simulação.

- **Função de Densidade de Probabilidade.** A função de densidade de probabilidade exibe a distribuição dos valores alvo, permitindo determinar a probabilidade de que o alvo esteja dentro de uma determinada região. Para metas com resultados definidos fixos, como "serviço ruim", "serviço justo", "serviço bom" e "serviço excelente" - é gerado um gráfico de barras que exibe a porcentagem de casos que se enquadram em cada categoria do alvo.
- **Função de Distribuição Cumulativa.** A função de distribuição acumulativa exibe a probabilidade de que o valor da resposta seja menor ou igual a um valor especificado.

**Gráficos de Tornado.** Os gráficos de tornado são gráficos de barras que exibem relacionamentos entre alvos e entradas simuladas usando uma variedade de métricas.

- **Correlação da resposta com a entrada.** Esta opção cria um gráfico tornado dos coeficientes de correlação entre um determinado alvo e cada uma de suas entradas simuladas.
- **Contribuição para variância.** Essa opção cria um gráfico de tornado que exibe a contribuição para a variação de um destino de cada uma de suas entradas simuladas, permitindo avaliar o grau em que cada entrada contribui para a incerteza geral no destino.
- **Sensibilidade de resposta para alterar.** Esta opção cria um gráfico de tornado que exibe o efeito no alvo de modular cada entrada simulada por mais ou menos um desvio padrão da distribuição associada à entrada.

**Gráficos de dispersão de respostas versus entradas.** Esta opção gera gráficos de dispersão de destinos versus entradas simuladas.

**Box plots de distribuições de resposta.** Esta opção gera gráficos de caixa das distribuições de destino.

**Tabela de quartis.** Esta opção gera uma tabela dos quartis das distribuições de destino. Os quartis de uma distribuição são os percentis 25, 50 e 75 da distribuição e dividem as observações em quatro grupos de igual tamanho.

**Correlações e tabela de contingência para entradas.** Esta opção exibe uma tabela de coeficientes de correlação entre as entradas simuladas. Uma tabela de contingência de associações entre entradas com uma distribuição categórica é exibida quando o plano de simulação especifica a geração de dados categóricos de uma tabela de contingência.

**Sobrepôr resultados a partir de respostas separadas.** Se o modelo preditivo que você está simulando contiver vários destinos, será possível especificar se os resultados de destinos separados serão exibidos em um único gráfico. Essa configuração se aplica a gráficos para funções de densidade de probabilidade, funções de distribuição cumulativa e gráficos de caixa. Por exemplo, se você selecionar esta opção, as funções de densidade de probabilidade para todos os alvos serão exibidas em um único gráfico.

**Salvar o plano para esta simulação.** É possível salvar quaisquer modificações em sua simulação em um arquivo de plano de simulação. Os arquivos de plano de simulação possuem a extensão *.splan*. É possível reabrir o plano na caixa de diálogo Executar simulação ou no Construtor de Simulação. Os planos de simulação incluem todas as especificações, exceto as configurações de saída.

**Salvar os dados simulados como um novo arquivo de dados.** É possível salvar as entradas simuladas, entradas fixas e valores de resposta preditos em um arquivo de dados do SPSS Statistics, em um novo conjunto de dados na sessão atual ou em um arquivo Excel. Cada caso (ou linha) do arquivo de dados consiste nos valores previstos dos alvos junto com as entradas simuladas e entradas fixas que geram os valores alvo. Quando a análise de sensibilidade é especificada, cada iteração dá origem a um conjunto contíguo de casos que são rotulados com o número da iteração.

Se você precisar de mais customização de saída do que a disponível aqui, considere executar sua simulação no Construtor de Simulação. Consulte o tópico [“Para executar uma simulação a partir de um plano de simulação” na página 250](#) para obter mais informações

## Trabalhando com saída de gráfico a partir da simulação

Vários gráficos gerados a partir de uma simulação possuem recursos interativos que permitem customizar a exibição. Os recursos interativos estão disponíveis ativando (clcando duas vezes) o objeto de gráfico no Visualizador de saída. Todos os gráficos de simulação são visualizações de gráfico. para obter mais informações.

**Gráficos de função de densidade de probabilidade para variáveis resposta contínuas.** Este gráfico tem duas linhas de referência verticais deslizantes que dividem o gráfico em regiões separadas. A tabela abaixo do gráfico exibe a probabilidade de que o alvo esteja em cada uma das regiões. Se várias funções de densidade forem exibidas no mesmo gráfico, a tabela terá uma linha separada para as probabilidades associadas a cada função de densidade. Cada uma das linhas de referência tem um controle deslizante (triângulo invertido) que permite mover facilmente a linha. Uma série de recursos adicionais estão disponíveis clicando no botão **Opções de gráfico** no gráfico. Em particular, é possível configurar explicitamente as posições dos controles deslizantes, adicionar linhas de referência fixas e alterar a visualização do gráfico de uma curva contínua para um histograma ou vice-versa. Consulte o tópico [“Opções de Gráfico” na página 265](#) para obter mais informações

**Gráficos de função de distribuição acumulativa para variáveis resposta contínuas.** Este gráfico tem as mesmas duas linhas de referência verticais móveis e a tabela associada descrita para o gráfico da função de densidade de probabilidade acima. Ele também fornece acesso à caixa de diálogo Opções do gráfico, que permite definir explicitamente as posições dos controles deslizantes, adicionar linhas de referência fixas e especificar se a função de distribuição cumulativa é exibida como uma função crescente (o padrão) ou uma função decrescente. Consulte o tópico [“Opções de Gráfico” na página 265](#) para obter mais informações



### **Gráficos de barras para variáveis respostas categóricas com iteração de análise de sensibilidade.**

Para destinos categóricos com iterações de análise de sensibilidade, os resultados da categoria de destino prevista são exibidos como um gráfico de barras agrupado que inclui os resultados de todas as iterações. O gráfico inclui uma lista suspensa que permite agrupar por categoria ou iteração. Para modelos de cluster de duas etapas e modelos de cluster K-médias, é possível optar por agrupar no número do cluster ou na iteração.

**Box plots para respostas múltiplas com iterações de análise de sensibilidade.** Para modelos preditivos com vários destinos contínuos e iterações de análise de sensibilidade, optar por exibir gráficos de caixa para todos os destinos em um único gráfico produz um gráfico de caixa agrupado. O gráfico inclui uma lista suspensa que permite agrupar no destino ou na iteração.

## **Opções de Gráfico**

O diálogo Opções de Gráfico permite customizar a exibição de gráficos ativados de funções de densidade de probabilidade e de funções de distribuição acumulativa gerados a partir de uma simulação.

**Visualizar.** A lista suspensa **Visualizar** se aplica apenas ao gráfico de função de densidade de probabilidade. Ela permite alternar a visualização do gráfico de uma curva contínua para um histograma. Este recurso não está disponível quando diversas funções de densidade são exibidas no mesmo gráfico. Nesse caso, as funções de densidade podem ser visualizadas somente como curvas contínuas.

**Ordem.** A lista suspensa **Ordem** se aplica somente ao gráfico de função de distribuição cumulativa. Ela especifica se a função de distribuição acumulativa é exibida como uma função ascendente (o padrão) ou uma função decrescente. Quando exibida como uma função decrescente, o valor da função em um determinado ponto no eixo horizontal é a probabilidade de que a resposta esteja à direita desse ponto.

**Posições da régua de controle.** É possível configurar explicitamente as posições das linhas de referência deslizantes ao inserir valores nas caixas de texto Superior e Inferior. É possível remover a linha à esquerda ao selecionar **-Infinito**, configurando efetivamente a posição para infinito negativo, e remover a linha à direita ao selecionar **Infinito**, configurando efetivamente sua posição para infinito.

**Linhas de referência.** É possível incluir várias linhas de referência verticais fixas nas funções de densidade de probabilidade e nas funções de distribuição acumulativa. Quando diversas funções são exibidas em um único gráfico (devido a diversas respostas ou resultados de iterações da análise de sensibilidade), é possível especificar as funções específicas às quais as linhas serão aplicadas.

- **Sigmas.** É possível incluir linhas de referência em mais e menos um número especificado de desvios padrão da média de uma resposta.
- **Percentis.** É possível incluir linhas de referência em um ou dois valores de percentil da distribuição de uma resposta ao inserir valores nas caixas de texto Inferior e Superior. Por exemplo, um valor de 95 na caixa de texto Superior representa o 95º percentil, que é o valor abaixo do qual 95% das observações caem. Da mesma forma, um valor de 5 na caixa de texto Inferior representa o 5º percentil, que é o valor abaixo do qual 5% das observações caem.
- **Posições customizadas.** É possível incluir linhas de referência com valores especificados ao longo do eixo horizontal.

**Rotular linhas de referência.** Essa opção controla se os rótulos são aplicados às linhas de referência selecionadas.

As linhas de referência são removidas ao desmarcar a opção associada no diálogo Opções de Gráfico e clicar em **Continuar**.

## **Modelagem geoespacial**

As técnicas de modelagem geoespacial são projetadas para descobrir padrões em dados que incluem um componente geoespacial (mapa). O Assistente de Modelagem Geoespacial fornece métodos para analisar dados geoespaciais com e sem um componente de tempo.

## **Encontre associações com base em dados de eventos e geoespaciais (Regras de associações geoespaciais)**

Usando regras de associação geoespacial, é possível localizar padrões nos dados com base nas propriedades espaciais e não espaciais. Por exemplo, é possível identificar padrões em dados de crimes por localização e atributos demográficos. A partir desses padrões, é possível criar regras que preveem onde determinados tipos de crimes provavelmente ocorrerão.

## **Faça previsões usando séries temporais e dados geoespaciais (previsão espacial-temporal)**

A previsão espacial temporal usa dados que contêm dados de localização, campos de entrada para previsão (preditores), um ou mais campos de tempo e um campo de destino. Cada local possui várias linhas nos dados que representam os valores de cada preditor e o destino em cada intervalo de tempo.

## **Usando o assistente de modelagem geoespacial**

1. A partir dos menus, escolha:

**Analisar > Modelagem espacial e temporal > Modelagem espacial**

2. Siga as etapas no assistente.

## **Exemplos**

Exemplos detalhados estão disponíveis no sistema de ajuda.

- Regras de associação geoespacial : **Ajuda > Tópicos > Estudos de Caso > Base de Estatísticas > Regras de associação espacial**
- Predição temporal espacial: **Ajuda > Tópicos > Estudos de Caso > Base de Estatísticas > Predição temporal espacial**

## **Selecionando Mapas**

A modelagem geoespacial pode utilizar uma ou mais origens de dados do mapa. As origens de dados do mapa contêm informações que definem áreas geográficas e outros recursos geográficos, como estradas ou rios. Muitas origens de mapas também contêm dados demográficos ou outros dados descritivos e dados de eventos, como relatórios de crimes ou taxas de desemprego. Você pode usar um arquivo de especificação de mapa definido anteriormente ou definir especificações de mapa aqui e salvar essas especificações para uso subsequente.

### **Carregar uma Especificação de Mapa**

Carrega um arquivo de especificação de mapa (.mplan) definido anteriormente. As origens de dados de mapa que você define aqui podem ser salvas em um arquivo de especificação de mapa. Para previsão espacial temporal, se você selecionar um arquivo de especificação de mapa que identifique mais de um mapa, você será solicitado a selecionar um mapa do arquivo.

### **Incluir Arquivo de Mapa**

Adicione um arquivo de formato ESRI (.shp) ou archive ZIP que contenha um arquivo de formato ESRI.

- Deve haver um arquivo .dbf correspondente no mesmo local que o arquivo .shp e esse arquivo deve ter o mesmo nome raiz do arquivo .shp.
- Se o arquivo for um archive ZIP, os arquivos .shp e .dbf deverão ter o mesmo nome raiz que o archive .zip. Se o arquivo for um arquivo .zip, os arquivos .shp e .dbf devem ter o mesmo nome raiz do arquivo .zip.
- Se não houver um arquivo de projeção (.prj) correspondente, você será solicitado a selecionar um sistema de projeção.

### **Relacionamento**

Para regras de associação geoespacial, esta coluna define como os eventos se relacionam com as feições no mapa. Essa configuração não está disponível para previsão temporal espacial.

## **Mover para Cima, Mover para baixo**

A ordem das camadas dos elementos do mapa é determinada pela ordem em que elas aparecem na lista. O primeiro mapa na lista é a camada inferior.

## **Selecionando um Mapa**

Para predição espacial temporal, se você selecionar um arquivo de especificação de mapa que identifique mais de um mapa, será solicitado que você selecione um mapa do arquivo. A predição espacial temporal não oferece suporte a vários mapas.

## **Relacionamento Geoespacial**

Para regras de associação geoespacial, a caixa de diálogo Relacionamento Geoespacial define como os eventos se relacionam com as feições no mapa.

- Essa configuração se aplica apenas a regras de associação geoespacial.
- Essa configuração afeta apenas as fontes de dados associadas a mapas especificados como dados de contexto na etapa de seleção de fontes de dados.

### **Relacionamento**

#### **Fechar**

O evento ocorre próximo a um ponto ou área especificada no mapa.

#### **Dentro de**

O evento ocorre dentro de uma área especificada no mapa.

#### **Contém**

A área de eventos contém um objeto de contexto de mapa.

#### **Faz intersecção com**

As localizações onde as linhas ou regiões de diferentes mapas se cruzam.

#### **Cruz**

Para vários mapas, locais em que as linhas (para estradas, rios, ferrovias) de diferentes mapas se cruzam.

#### **Norte de, Sul de, Leste de, Oeste de**

O evento ocorre dentro de uma área ao norte, sul, leste ou oeste de um ponto especificado no mapa.

## **Configurar Sistema de Coordenadas**

Se não houver um arquivo de projeção (.prj) com o mapa ou você definir dois campos de uma fonte de dados como um conjunto de coordenadas, você deve definir o sistema de coordenadas.

### **Geográfico padrão (longitude e latitude)**

O sistema de coordenadas é longitude e latitude.

### **Cartesiano Simples (X e Y)**

O sistema de coordenadas são coordenadas X e Y simples.

### **Usar um ID bem conhecido (WKID)**

"ID bem conhecido" para projeções comuns.

### **Usar um nome de sistema de coordenadas**

O sistema de coordenadas baseia-se na projeção nomeada. O nome é colocado entre parênteses.

## **Configurando a Projeção**

Se o sistema de projeção não puder ser determinado a partir das informações fornecidas com o mapa, será necessário especificar o sistema de projeção. A causa mais comum dessa condição é a ausência de um arquivo de projeção (.prj) associado ao mapa ou um arquivo de projeção que não pode ser usado.

- **Uma cidade, região ou país (Mercator)**
- **Um país grande, vários países ou continentes (Winkel Tripel)**

- **Uma área muito próxima ao equador (Mercator)**
- **Uma área próxima a um dos polos (Estereográfico)**

A projeção Mercator é uma projeção comum utilizada em muitos mapas. Essa projeção trata o globo como um cilindro que se desenrola sobre uma superfície plana. A projeção de Mercator distorce o tamanho e a forma de objetos grandes. Essa distorção aumenta à medida que você se afasta do equador e se aproxima dos pólos. As projeções Winkel Tripel e Stereographic fazem ajustes para o fato de um mapa representar uma porção de uma esfera tridimensional exibida em duas dimensões.

## Sistema de Projeção e de Coordenadas

Se você selecionar mais de um mapa e os mapas possuírem sistemas de projeção e de coordenadas diferentes, o mapa deverá ser selecionado com o sistema de projeção que deseja utilizar. Esse sistema de projeção será usado para todos os mapas quando eles forem combinados em uma saída.

## Origens de Dados

Uma origem de dados pode ser um arquivo dBase que é fornecido com o arquivo de formas, com um arquivo de dados do IBM SPSS Statistics ou com um conjunto de dados aberto na sessão atual.

**Dados de Contexto.** Os dados de contexto identificam recursos no mapa. Um dado de contexto também pode conter campos que podem ser usados como entradas para o modelo. Para usar um arquivo dBase de contexto (.dbf) associado a um arquivo de forma de mapa (.shp), o arquivo dBase de contexto deve estar no mesmo local que o arquivo de forma e deve ter o mesmo nome raiz. Por exemplo, se o arquivo de formas for geodata.shp, o arquivo dBase deverá ser denominado geodata.dbf

**Dados de Eventos.** Os dados de eventos contêm informações sobre eventos que ocorrem, como crimes ou acidentes. Esta opção está disponível apenas para regras de associação geoespaciais.

**Densidade de Ponto.** Intervalo de tempo e dados coordenados para estimativas de densidade do kernel. Esta opção está disponível apenas para predição espacial temporal.

**Incluir.** Abre um diálogo para incluir origens de dados. Uma origem de dados pode ser um arquivo dBase que é fornecido com o arquivo de formas, com um arquivo de dados do IBM SPSS Statistics ou com um conjunto de dados aberto na sessão atual.

**Associar.** Abre uma caixa de diálogo para especificar os identificadores (coordenadas ou chaves) usados para associar dados a mapas. Cada origem de dados deve conter um ou mais identificadores que associam os dados ao mapa. Os arquivos dBase que vêm com um arquivo shape normalmente contêm um campo que é usado automaticamente como o identificador padrão. Para outras origens de dados, deve-se especificar os campos que são usados como identificadores.

**Validar Chave.** Abre uma caixa de diálogo para validar a correspondência de chave entre o mapa e a origem de dados.

## Regras de associação geoespacial

- Pelo menos uma origem de dados deve ser uma origem de dados de eventos.
- Todas as fontes de dados de eventos devem usar a mesma forma de identificadores de associação de mapa: coordenadas ou valores de chave.
- Se as origens de dados de eventos estiverem associadas aos mapas com valores de chave, todas as origens de eventos deverão usar o mesmo tipo de recurso de mapa (por exemplo, polígonos, pontos, linhas).

## Predição temporal espacial

- Deve haver uma origem de dados de contexto.
- Se houver apenas uma origem de dados (um arquivo de dados sem mapa associado), ela deverá incluir valores de coordenadas.

- Se você tiver duas fontes de dados, uma fonte de dados deve ser dados de contexto e a outra fonte de dados deve ser dados de densidade de ponto.
- Não é possível incluir mais de duas origens de dados.

## Incluir uma Origem de Dados

Uma origem de dados pode ser um arquivo dBase que é fornecido com o arquivo de formas ou arquivo de contexto, com um arquivo de dados do IBM SPSS Statistics ou com um conjunto de dados aberto na sessão atual.

É possível incluir a mesma origem de dados várias vezes se quiser usar uma associação espacial diferente com cada uma.

## Associação de Dados e Mapa

Cada origem de dados deve conter um ou mais identificadores que associam os dados ao mapa.

### Coordenadas

A origem de dados contém campos que representam as coordenadas cartesianas, selecione os campos que representam as coordenadas X e Y. Para regras de associação geoespacial, também pode haver uma coordenada Z.

### Valores da chave

Os valores de chave nos campos na fonte de dados correspondem às chaves de mapa selecionadas. Por exemplo, um mapa de regiões pode ter um identificador de nome (chave de mapa) rotulando cada região. Esse identificador corresponde a um campo nos dados que também contém os nomes das regiões (chave de dados). Os campos são combinados com as chaves do mapa com base na ordem em que são exibidos nas duas listas.

## Validar chaves

A caixa de diálogo Validar Chaves fornece um resumo da correspondência de registros entre o mapa e a fonte de dados, com base nas chaves de identificador selecionadas. Se houver valores de chave de dados sem correspondência, será possível combiná-los manualmente para mapear valores de chave.

## Regras de associação geoespacial

Para regras de associação geoespaciais, depois de definir mapas e origens de dados, as etapas restantes no assistente são:

- Se houver diversas origens de dados do evento, defina como as origens de dados do evento são mescladas.
- Selecione os campos a serem usados como condições e predições na análise.

Opcionalmente, é possível:

- Selecione diferentes opções de saída.
- Salve um arquivo de modelo de pontuação.
- Crie novos campos para valores previstos e regras nas fontes de dados usadas no modelo.
- Customize as configurações para criar regras de associação.
- Customize as configurações de categorização e agregação.

## Definir Campos de Dados do Evento

Para regras de associação geoespacial, se houver mais de uma origem de dados do evento, as origens de dados do evento serão mescladas.

- Por padrão, apenas os campos que forem comuns a todas as origens de dados do evento são incluídos.
- É possível exibir uma lista de campos comuns, campos para uma origem de dados específica ou campos a partir de todas as origens de dados e selecionar os campos que você deseja incluir.

- Para campos comuns, o **Tipo** e a **Medição** devem ser os mesmos para todas as origens de dados. Se houver conflitos, será possível especificar o tipo e o nível de medição para uso em cada campo comum.

## Selecionar Campos

A lista de campos disponíveis inclui campos das origens de dados de eventos e campos das origens de dados de contexto.

- É possível controlar a lista de campos exibidos ao selecionar uma origem de dados a partir da lista **Origens de Dados**.
- Deve-se selecionar pelo menos dois campos. Pelo menos um deve ser uma condição e pelo menos um deve ser uma predição. Existem diversas formas para atender a esse requisito, incluindo selecionar dois campos para a lista **Ambos (Condição e Predição)**.
- As regras de associação preveem valores dos campos de previsão com base nos valores dos campos de condição. Por exemplo, na regra "Se  $x=1$  e  $y=2$ , então  $z=3$ ", os valores de  $x$  e  $y$  são condições e o valor de  $z$  é a predição.

## Saída

### Tabelas de Regras

Cada tabela de regras exibe as principais regras e valores para confiança, suporte a regras, elevação, suporte a condições e capacidade de implementabilidade. Cada tabela é armazenada pelos valores dos critérios selecionados. É possível exibir todas as regras ou o **Número** de regras principais, com base no critério selecionado.

### Nuvem de palavras classificáveis

Uma lista de regras principais com base nos valores do critério selecionado. O tamanho do texto indica a importância relativa da regra. O objeto de saída interativa contém as principais regras para confiança, suporte a regras, elevação, suporte a condições e capacidade de implementabilidade. O critério selecionado determina qual lista de regras é exibida por padrão. É possível selecionar um critério diferente de forma interativa na saída. **Máximo de regras para exibir** determina o número de regras que são exibidas na saída.

### Mapa

Gráfico de barras e mapa interativos das regras principais, com base no critério selecionado. Cada objeto de saída interativa contém as principais regras para confiança, suporte a regras, elevação, suporte a condições e capacidade de implementabilidade. O critério selecionado determina qual lista de regras é exibida por padrão. É possível selecionar um critério diferente de forma interativa na saída. **Máximo de regras para exibir** determina o número de regras que são exibidas na saída.

### Tabelas de informações de modelo

#### Transformações de campo.

Descreve as transformações que são aplicadas aos campos usados na análise.

#### Resumo do registro.

Número e percentual de registros incluídos e excluídos.

#### Estatísticas da regra.

Estatísticas básicas do suporte de condição, confiança, suporte de regra, elevação e implementabilidade. As estatísticas incluem média, mínimo, máximo e desvio padrão.

#### Itens Mais Frequentes.

Itens que ocorrem com mais frequência. Um item está incluído em uma condição ou predição em uma regra. Por exemplo,  $age < 18$  or  $gender=female$ .

#### Campos Mais Frequentes.

Campos que ocorrem com mais frequência nas regras.

#### Entradas excluídas.

Campos excluídos da análise e o motivo da exclusão de cada campo.

## **Critério para tabelas de regras, nuvem de palavras e mapas**

### **Confiança.**

A porcentagem de predições de regra correta.

### **Suporte de regra.**

A porcentagem de casos para os quais a regra é verdadeira. Por exemplo, se a regra for "Se  $x=1$  e  $y=2$ , então  $z=3$ ", o suporte da regra será a porcentagem real de casos nos dados para os quais  $x=1$ ,  $y=2$  e  $z=3$ .

### **Elevação.**

Elevação é uma medida do quanto a regra melhora a predição em comparação com a chance aleatória. É a razão entre as previsões corretas e a ocorrência geral do valor predito. O valor deve ser maior que 1. Por exemplo, se o valor previsto ocorrer 20% das vezes e a confiança na previsão for 80%, o valor de elevação será 4.

### **Suporte de condição.**

A porcentagem de casos para os quais a condição da regra existe. Por exemplo, se a regra for "Se  $x=1$  e  $y=2$ , então  $z=3$ ", o suporte à condição será a proporção de casos nos dados para os quais  $x=1$  e  $y=2$ .

### **Implementabilidade.**

A porcentagem de predições incorretas quando as condições são verdadeiras. A implementabilidade é igual a (1 confiança) multiplicado pelo suporte à condição ou suporte à condição menos o suporte à regra.

## **Salve**

### **Salvar o mapa e dados de contexto como uma especificação de mapa**

Salve as especificações de mapa em um arquivo externo (.mplan). É possível carregar este arquivo de especificação de mapa no assistente para análises subsequentes. Também é possível utilizar o arquivo de especificação de mapa com o comando SPATIAL ASSOCIATION RULES.

### **Copie qualquer mapa e arquivos de dados na especificação**

Dados de arquivos de forma de mapa, arquivos de dados externos e conjuntos de dados usados na especificação de mapa são salvos no arquivo de especificação de mapa.

### **Escoragem**

Salva os melhores valores de regra, valores de confiança para as regras e valores de ID numérico para as regras como novos campos na fonte de dados especificada.

### **Origem de Dados para Escorar**

Uma ou mais origens de dados em que os novos campos são criados. Se a origem de dados não estiver aberta na sessão atual, ela será aberta na sessão atual. Deve-se salvar explicitamente o arquivo modificado para salvar os novos campos.

### **Valores de destino**

Crie novos campos para os campos de destino (predição) selecionados.

- Dois novos campos são criados para cada campo de destino: valor previsto e valor de confiança.
- Para campos de destino contínuos (escala), o valor previsto é uma sequência que descreve um intervalo de valores. Um valor no formato "(valor1, valor2]" significa "maior que valor1 e menor ou igual a valor2".

### **Número de melhores regras**

Crie novos campos para o número de melhores regras especificadas. Três novos campos são criados para cada regra: valor da regra, valor de confiança e um valor de ID numérico para a regra.

### **Prefixo do nome**

Prefixo a ser utilizado para os novos nomes de campo.

## **Construção de Regra**

Os parâmetros de construção de regra configuram os critérios para as regras de associação geradas.

### **Itens por Regra**

Número de valores de campo que podem ser incluídos nas condições da regra e predições. O número total de itens não pode exceder 10. Por exemplo, na regra "If x=1 and y=2, then z=3", há dois itens de condição e um item de predição.

#### **Máximo de predições.**

Número máximo de valores de campo que podem ocorrer nas predições para uma regra.

#### **Máximo de condições.**

Número máximo de valores de campo que podem ocorrer nas condições para uma regra.

### **Excluir par**

Exclui os pares de campos especificados de serem incluídos na mesma regra.

### **Critérios de Regra**

#### **Confiança.**

Confiança mínima que uma regra deve ter para ser incluída na saída. A confiança é a porcentagem de predições corretas.

#### **Suporte de regra.**

Suporte de regra mínimo que uma regra deve ter para ser incluída na saída. O valor representa a porcentagem de casos para os quais a regra é verdadeira nos dados observados. Por exemplo, se a regra for "If x=1 and y=2, then z=3," o suporte de regra será a porcentagem real de casos nos dados para os quais x=1, y=2 e z=3.

#### **Suporte de condição.**

Suporte de condição mínimo que uma regra deve ter para ser incluída na saída. O valor representa a porcentagem de casos para os quais a condição existe. Por exemplo, se a regra for "If x=1 and y=2, then z=3", o suporte de condição é a porcentagem de casos nos dados para os quais x=1 e y=2.

#### **Elevação.**

Elevação mínima que uma regra deve ter para ser incluída na saída. Elevação é uma medida do quanto a regra melhora a predição em uma chance aleatória. É a razão de predições corretas para a ocorrência geral do valor predito. Por exemplo, se o valor predito ocorrer 20% do tempo e a confiança na predição for 80%, então o valor de elevação será 4.

### **Tratar como iguais**

Identifica pares de campos que devem ser tratados como o mesmo campo.

## **Categorização e Agregação**

- A agregação é necessária quando há mais registros nos dados do que feições no mapa. Por exemplo, você possui registros de dados para estados individuais, mas também possui um mapa de estados.
- É possível especificar o método de medida de sumarização agregada para campos contínuos e ordinais. Os campos nominais são agregados com base no valor modal.

### **Contínuo**

Para campos contínuos (escala), a medida de sumarização pode ser média, mediana ou soma.

### **Ordinal**

Para campos ordinais, a medida de resumo pode ser mediana, modo, mais alta ou mais baixa.

### **Número de categorias**

Configura o número máximo de categorias para campos contínuos (escala). Os campos contínuos são sempre agrupados ou "agrupados" em intervalos de valores. Por exemplo: menor ou igual a 5, maior que 5 e menor ou igual a 10, ou maior que 10.

### **Agregar o mapa**

Aplicar agregação para dados e mapas.

### **Configurações customizadas para campos específicos**

É possível substituir a medida de sumarização padrão e o número de categorias para campos específicos.

- Clique no ícone para abrir o diálogo **Seletor de Campos** e selecione um campo para incluir na lista.



- Na coluna **Agregação**, selecione uma medida de sumarização.
- Para campos contínuos, clique no botão na coluna **Categorias** para especificar um número customizado de categorias para o campo no diálogo **Categorias**.

## Predição temporal espacial

Para predição temporal espacial, após definir mapas e origens de dados, as etapas restantes no assistente são:

- Especifique o campo de destino, campos de tempo e preditores opcionais.
- Defina intervalos de tempo ou períodos cíclicos para campos de tempo.

Opcionalmente, é possível:

- Selecione diferentes opções de saída.
- Customize os parâmetros de construção do modelo.
- Customize as configurações de agregação.
- Salvar valores preditos em um conjunto de dados na sessão atual ou em um arquivo de dados formatados do IBM SPSS Statistics.

## Selecionar Campos

A lista de campos disponíveis inclui campos de origens de dados selecionadas. É possível controlar a lista de campos exibidos ao selecionar uma origem de dados a partir da lista **Origens de Dados**.

### Target

Um campo de destino é necessário. O destino é o campo para o qual os valores são preditos.

- O campo de destino deve ser um campo numérico contínuo (escala).
- Se houver duas origens de dados, o destino será as estimativas de densidade do kernel e "Densidade" será exibido como o nome do destino. Essa seleção não pode ser alterada.

### Preditores

Um ou mais campos preditores podem ser especificados. Essa configuração é opcional.

### Campos de tempo

Deve-se selecionar um ou mais campos que representem períodos de tempo ou selecionar **Períodos Cíclicos**.

- Se houver duas origens de dados, deve-se selecionar os campos de tempo para ambas as origens de dados. Ambos os campos de tempo devem representar o mesmo intervalo.
- Para períodos cíclicos, deve-se especificar os campos que definem ciclos de periodicidade no painel **Intervalos de Tempo** do assistente.

## Intervalos de Tempo

As opções nesse painel baseiam-se na opção de **Campos de Tempo** ou **Período cíclico** no passo para selecionar campos.

### Campos de Hora

**Campos de Tempo Selecionados.** Se você selecionar um ou mais campos de tempo na etapa de seleção de campos, esses campos serão exibidos nesta lista.

**Intervalo de Tempo.** Selecione o intervalo de tempo apropriado na lista. Dependendo do intervalo de tempo, você também pode especificar outras configurações, como intervalo entre observações (incremento) e valor inicial. Este intervalo de tempo é usado para todos os campos de tempo selecionados.

- O procedimento assume que todos os casos (registros) representam intervalos igualmente espaçados.

- Com base no intervalo de tempo selecionado, o procedimento pode detectar observações ausentes ou várias observações no mesmo intervalo de tempo que precisam ser agregadas. Por exemplo, se o intervalo de tempo for dias e a data 27/10/2014 for seguida por 29/10/2014, haverá uma observação ausente para 28/10/2014. Se o intervalo de tempo for mês, várias datas no mesmo mês serão agregadas.
- Para alguns intervalos de tempo, a configuração adicional pode definir intervalos nos intervalos normais igualmente espaçados. Por exemplo, se o intervalo de tempo for dias, mas apenas os dias da semana forem válidos, você poderá especificar que há cinco dias em uma semana e a semana começa na segunda-feira.
- Se o campo de tempo selecionado não for um campo de formato de data ou de formato de hora, o intervalo de tempo será configurado automaticamente para **Períodos** e não poderá ser alterado.

## Campos de ciclo

Se selecionar **Período cíclico** no passo para selecionar campos, deve-se especificar os campos que definem os períodos cíclicos. Um período cíclico identifica variação cíclica repetitiva, como o número de meses em um ano ou o número de dias em uma semana.

- É possível especificar até três campos que definem períodos cíclicos.
- O primeiro campo de ciclo representa o nível mais alto do ciclo. Por exemplo, se houver variação cíclica por ano, trimestre e mês, o campo que representa o ano é o campo do primeiro ciclo.
- A duração do ciclo para os campos do primeiro e do segundo ciclo é a periodicidade no nível subsequente. Por exemplo, se os campos de ciclo forem ano, trimestre e mês, a duração do primeiro ciclo será 4 e a duração do segundo ciclo será 3.
- O valor inicial para os campos do segundo e terceiro ciclo é o primeiro valor em cada um desses períodos cíclicos.
- A duração do ciclo e os valores iniciais devem ser números inteiros positivos.

## Agregação

- Se selecionar quaisquer **Preditores** no passo para selecionar campos, será possível selecionar o método de sumarização de agregação para os preditores.
- A agregação é necessária quando houver mais de um registro em um intervalo de tempo definido. Por exemplo, se o intervalo de tempo for mês, várias datas no mesmo mês serão agregadas.
- É possível especificar o método de medida de resumo de agregação para campos contínuos e ordinais. Os campos nominais são agregados com base no valor modal.

### Contínuo

Para campos contínuos (escala), a medida de sumarização pode ser média, mediana ou soma.

### Ordinal

Para campos ordinais, a medida de resumo pode ser mediana, modo, mais alta ou mais baixa.

### Configurações customizadas para campos específicos

É possível substituir a medida de sumarização de agregação padrão para preditores específicos.

- Clique no ícone para abrir o diálogo **Seletor de Campos** e selecione um campo para incluir na lista.
- Na coluna **Agregação**, selecione uma medida de sumarização.

## Saída

### Mapas

#### Valores de destino.

Mapa de valores para o campo de destino selecionado.

#### Correlação

Mapa de correlações.

## Clusters

Mapa que destaca os clusters de localizações que são semelhantes entre si. Mapas de clusters estão disponíveis apenas para modelos empíricos.

### Limite de similaridade de localização

A similaridade que é necessária para criar os clusters. O valor deve ser um número maior que zero e menor que 1.

### Especifique o número máximo de clusters.

O número máximo de clusters para exibir.

## Tabelas de avaliação de modelo

### Especificações de modelo.

Resumo das especificações que são usadas para executar a análise, incluindo campos de destino, entrada e local.

### Resumo de informações temporais.

Identifica os campos e intervalos de tempo que são utilizados no modelo.

### Teste de efeitos na estrutura média.

A saída inclui o valor da estatística de teste, graus de liberdade e nível de significância para o modelo e cada efeito.

### Estrutura média dos coeficientes do modelo.

A saída inclui o valor do coeficiente, erro padrão, valor da estatística de teste, nível de significância e intervalos de confiança para cada termo do modelo.

### Coefficientes autorregressivos.

A saída inclui o valor do coeficiente, o erro padrão, o valor da estatística de teste, o nível de significância e os intervalos de confiança para cada defasagem.

### Testes de Covariância Espacial.

Para modelos paramétricos baseados em variogramas, exibe a qualidade dos resultados do teste de ajuste para a estrutura de covariância espacial. Os resultados do teste podem determinar se a estrutura de covariância espacial deve ser modelada parametricamente ou usar um modelo não paramétrico.

### Covariância espacial paramétrica.

Para modelos paramétricos baseados em variogramas, exibe estimativas de parâmetros para covariância espacial paramétrica.

## Opções de Modelo

### Configurações de modelo

#### Incluir automaticamente um intercepto

Inclua a interceptação no modelo.

#### Atraso máximo de autoregressão

O lag máximo de autorregressão. O valor deve ser um número inteiro entre 1 e 5.

### Covariância espacial

Especifica o método de estimativa para covariância espacial.

#### Paramétrico

O método de estimação é paramétrico. O método pode ser **Gaussiano**, **Exponencial** ou **Potência Exponencial**. Para Potência Exponencial, é possível especificar o valor de **Potência**.

#### Não paramétrico

O método de estimação é não paramétrico.

## Salvar

### Salvar o mapa e dados de contexto como uma especificação de mapa

Salve as especificações de mapa em um arquivo externo (.mplan). É possível carregar esse arquivo de especificação de mapa no assistente para análises subsequentes. Também é possível utilizar o arquivo de especificação de mapa com o comando SPATIAL TEMPORAL PREDICTION.

### **Copiar quaisquer arquivos de mapas e de dados na especificação**

Dados de arquivos de formas de mapa, arquivos de dados externos e conjuntos de dados que são utilizados na especificação de mapa são salvos no arquivo de especificação de mapa.

### **Pontuação**

Salva valores preditos, a variância e limites de confiança superior e inferior para o campo de destino no arquivo de dados selecionado.

- É possível salvar valores preditos em um conjunto de dados aberto na sessão atual ou em um arquivo de dados de formato do IBM SPSS Statistics.
- O arquivo de dados não pode ser uma origem de dados que é utilizada no modelo.
- O arquivo de dados deve conter todos os campos de tempo e preditores que são utilizados no modelo.
- Os valores de tempo devem ser maiores que os valores de tempo utilizados no modelo.

### **Avançado**

#### **Máximo de casos com valores ausentes (%)**

A porcentagem máxima de casos com valores ausentes.

#### **Nível de significância**

O nível de significância para determinar se um modelo paramétrico baseado em variograma é apropriado. O valor deve ser maior do que 0 e menor do que 1. O valor padrão é 0,05. O nível de significância é utilizado no teste de qualidade de ajuste para estrutura de covariância espacial. A estatística de qualidade de ajuste é usada para determinar se deve ser usado um modelo paramétrico ou não paramétrico.

#### **Fator de incerteza (%)**

O fator de incerteza é um valor percentual que representa o crescimento da incerteza para previsões futuras. Os limites superior e inferior da incerteza de previsão aumentam na porcentagem especificada para cada etapa no futuro.

### **Conclusão**

Na última etapa do Assistente de Modelagem Geoespacial, é possível executar o modelo ou colar a sintaxe de comando gerada em uma janela de sintaxe. É possível modificar e salvar a sintaxe gerada para uso posterior.

## Avisos

---

Estas informações foram desenvolvidas para produtos e serviços oferecidos nos EUA. Esse material pode estar disponível a partir da IBM em outros idiomas. Entretanto, pode ser necessário que possua uma cópia do produto ou versão de produto nesse idioma a fim de acessá-lo.

É possível que a IBM não ofereça os produtos, serviços ou recursos discutidos nesta publicação em outros países. Consulte um representante IBM local para obter informações sobre produtos e serviços disponíveis atualmente em sua área. Qualquer referência a um produto, programa ou serviço IBM não está destinado a declarar ou implicar que apenas esse produto, programa ou serviço IBM possa ser usado. 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. É possível enviar consultas sobre licenças, por escrito, para:

*Gerência de Relações Comerciais e Industriais da IBM Brasil*

*Av. Pasteur, 138-146, Botafogo  
Botafogo  
Rio de Janeiro, RJCEP 22290-240*

Para consultas sobre 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 consultas sobre licença, por escrito, para:

*Intellectual Property Licensing*

*Legal and Intellectual Property Law  
IBM Japan Ltd.*

*19-21, Nihonbashi-Hakozakicho, Chuo-kuTokyo 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 A ELAS NÃO SE LIMITANDO, AS GARANTIAS IMPLÍCITAS DE NÃO INFRAÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO PROPÓSITO. Alguns países não permitem a exclusão de garantias expressas ou implícitas em certas transações; portanto, essa disposição pode não se aplicar ao Cliente.

Essas informações podem conter imprecisões técnicas ou erros tipográficos. Periodicamente, são feitas mudanças nas informações aqui contidas; tais mudanças serão incorporadas em novas edições da publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar os produtos e/ou programas 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 desse produto IBM e a utilização 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 este assunto com objetivo de permitir: (i) a troca de informações entre programas criados independentemente e outros programas (incluindo este) e (ii) a utilização mútua das informações trocadas, devem entrar em contato com:

*Av. Pasteur, 138-146, Botafogo  
Botafogo  
Rio de Janeiro, RJCEP 22290-240*

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. Os resultados de desempenho reais podem variar dependendo das configurações específicas e 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.

Essas informações contêm exemplos de dados e relatórios utilizados em operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos incluem nomes de indivíduos, 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 exemplo 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 nenhum 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:

© Copyright IBM Corp. 2021. Partes deste código são derivadas de Programas de Amostra da IBM Corp. Programas de amostra.

© Copyright IBM Corp. 1989 - 2021. Todos os direitos reservados.

## Marcas comerciais

---

IBM, o logotipo IBM e [ibm.com](http://ibm.com) são marcas comerciais ou marcas registradas da International Business Machines Corp., registradas em várias jurisdições no mundo inteiro. Outros nomes de produtos e serviços podem ser marcas registradas 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](http://www.ibm.com/legal/copytrade.shtml).

Adobe, o logotipo Adobe, PostScript e o logotipo PostScript são marcas ou marcas registradas 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 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 de suas afiliadas.





# Índice remissivo

## Caracteres Especiais

- em cubos OLAP [104](#)
- em Descritivos [89](#)
- em Estatísticas de razão [236](#), [238](#)
- em Frequências [87](#)
- em Médias [101](#)
- em Sumarizar [99](#)
- índice de concentração
  - em Estatísticas de razão [236](#), [238](#)
- Ômega de McDonald
  - na Análise de confiabilidade [227](#), [228](#)
- últimos(as)
  - em cubos OLAP [104](#)
  - em Médias [101](#)
  - em Sumarizar [99](#)

## A

- Agresti-Caffo
  - em Proporções de amostras independentes [113](#)
- Agresti-Coull
  - em proporções de uma amostra [107](#)
- Agresti-Min
  - em proporções de amostras emparelhadas [110](#)
- ajuste de distribuição
  - na simulação [254](#)
- ajuste de distribuição automática
  - na simulação [254](#)
- Ajuste de escala multidimensional
  - ajustando a escala de modelos [234](#)
  - condicionalidade [234](#)
  - criando matrizes de distância [234](#)
  - critérios [235](#)
  - definição de forma de dados [234](#)
  - dimensões [234](#)
  - estatísticas [233](#)
  - exemplo [233](#)
  - medidas de distância [234](#)
  - níveis de medida [234](#)
  - opções de exibição [235](#)
  - recursos adicionais do comando [235](#)
  - valores de transformação [234](#)
- Alfa de Cronbach
  - na Análise de confiabilidade [227](#)
- alocação de memória
  - na Análise de Cluster TwoStep [187](#)
- alpha de Cronbach
  - na Análise de confiabilidade [228](#)
- amostra de treinamento
  - na análise do vizinho mais próximo [172](#)
- amostra de validação
  - na análise do vizinho mais próximo [172](#)
- amostras relacionadas [215](#), [217](#)
- amplitude múltipla de Ryan-Einot-Gabriel-Welsch (*continuação*)
  - no One-Way ANOVA [120](#)
- Análise de Característica de Operação do Receptor
  - estatísticas e gráficos [246](#)
- análise de cluster
  - Análise de cluster hierárquica [193](#)
  - Análise de Cluster por K-Médias [195](#)
  - eficiência [196](#)
- Análise de cluster hierárquica
  - armazenando casos em cluster [193](#)
  - armazenando variáveis em cluster [193](#)
  - associação de cluster [195](#)
  - dendrogramas [195](#)
  - estatísticas [193](#)
  - exemplo [193](#)
  - gráfico icicle [195](#)
  - medidas de distância [194](#)
  - medidas de similaridade [194](#)
  - medidas de transformação [194](#)
  - métodos de clusterização [194](#)
  - orientação do gráfico [195](#)
  - recursos adicionais do comando [195](#)
  - salvando novas variáveis [195](#)
  - valores de transformação [194](#)
- Análise de cluster hierárquico
  - associação de cluster [194](#)
  - estatísticas [194](#)
  - matrizes de distância [194](#)
  - planejamentos de aglomeração [194](#)
- Análise de Cluster por K-Médias
  - associação de cluster [197](#)
  - critérios de convergência [196](#)
  - distâncias de cluster [197](#)
  - eficiência [196](#)
  - estatísticas [195](#), [197](#)
  - exemplos [195](#)
  - iterações [196](#)
  - métodos [195](#)
  - recursos adicionais do comando [197](#)
  - salvando informações do cluster [197](#)
  - valores omissos [197](#)
  - visão geral [195](#)
- Análise de cluster TwoStep [185](#)
- Análise de Cluster TwoStep
  - estatísticas [188](#)
  - opções [187](#)
  - salvar em arquivo externo [188](#)
  - salvar no arquivo de trabalho [188](#)
- análise de componentes principais [181](#), [182](#)
- Análise de Confiabilidade
  - coeficiente de correlação intraclasse [228](#)
  - correlações e covariâncias entre itens [228](#)
  - descritivas [228](#)
  - estatísticas [227](#), [228](#)
  - exemplo [227](#)
  - Kuder-Richardson 20 [228](#)
  - recursos adicionais do comando [230](#)

- Análise de Confiabilidade (*continuação*)
    - T 2 de Hotelling [228](#)
    - Tabela de ANOVA [228](#)
    - Teste de aditividade de Tukey [228](#)
  - análise de múltiplas respostas
    - Frequências de múltiplas respostas [219](#)
    - tabelas de frequência [219](#)
    - tabulação cruzada [220](#)
    - Tabulações cruzadas de múltiplas respostas [220](#)
  - Análise de potência
    - Estatísticas [1](#)
  - análise de sensibilidade
    - na simulação [257](#)
  - análise de série temporal
    - casos de previsão [166](#)
    - previsão [166](#)
  - análise de variância
    - em Médias [101](#)
    - na Estimativa de curva [165](#)
    - Na Regressão Linear [150](#)
    - no One-Way ANOVA [119](#)
  - Análise discriminante
    - coeficientes de função [179](#)
    - critérios [179](#)
    - definindo intervalos [178](#)
    - Distância de Mahalanobis [179](#)
    - estatística descritiva [179](#)
    - estatísticas [177](#), [179](#)
    - exemplo [177](#)
    - exportando informações de modelo [181](#)
    - gráficos [180](#)
    - Lambda de Wilks [179](#)
    - matriz de covariâncias [180](#)
    - matrizes [179](#)
    - métodos discriminantes [179](#)
    - métodos stepwise [177](#)
    - opções de exibição [179](#), [180](#)
    - probabilidades anteriores [180](#)
    - recursos adicionais do comando [181](#)
    - salvando variáveis de classificação [181](#)
    - selecionando casos [178](#)
    - V de Rao [179](#)
    - valores omissos [180](#)
    - variáveis de agrupamento [177](#)
    - variáveis independentes [177](#)
  - Análise do vizinho mais próximo
    - options [174](#)
    - partições [172](#)
    - saída [174](#)
    - salvando variáveis [173](#)
    - seleção de variável [172](#)
    - visualização do modelo [174](#)
    - vizinhos [171](#)
  - Análise fatorial
    - carregando gráficos [183](#)
    - convergência [182](#), [183](#)
    - descritiva [182](#)
    - escores dos fatores [184](#)
    - estatísticas [181](#), [182](#)
    - exemplo [181](#)
    - formato de exibição de coeficiente [184](#)
    - métodos de extração [182](#)
    - métodos de rotação [183](#)
    - recursos adicionais do comando [185](#)
  - Análise fatorial (*continuação*)
    - selecionando casos [182](#)
    - valores omissos [184](#)
    - visão geral [181](#)
  - Análise ROC
    - estatísticas e gráficos [245](#)
  - análise what-if
    - na simulação [257](#)
  - ANOVA
    - em Médias [101](#)
    - em modelos lineares [145](#)
    - modelo [125](#)
    - no GLM Univariate [123](#)
    - no One-Way ANOVA [119](#)
  - ANOVA de um fator
    - contrastes [120](#)
    - contrastes polinomiais [120](#)
    - estatísticas [122](#)
    - múltiplas comparações [120](#)
    - opções [122](#)
    - recursos adicionais do comando [123](#)
    - testes post hoc [120](#)
    - valores omissos [122](#)
    - variáveis de fator [119](#)
  - Anscombe
    - em proporções de uma amostra [107](#)
  - armazenamento em cluster
    - escolhendo um procedimento [185](#)
    - exibição geral [189](#)
    - visualizando clusters [189](#)
  - assimetria
    - em cubos OLAP [104](#)
    - em Descritivos [89](#)
    - em Frequências [87](#)
    - em Médias [101](#)
    - em Resumos de Relatórios em Colunas [225](#)
    - em sumarizações de relatórios em linhas [223](#)
    - em Sumarizar [99](#)
    - no Explore [92](#)
  - associação linear por linear
    - em Tabulações Cruzadas [95](#)
  - autovalores
    - na Análise fatorial [182](#)
    - Na Regressão Linear [150](#)
- ## B
- bagging
    - em modelos lineares [141](#)
  - Binário de Metanálise
    - Diálogo Análise [58](#)
    - Diálogo Aparar e Preencher [61](#)
    - Diálogo Contraste [60](#)
    - Diálogo Critérios [56](#)
    - Diálogo Gráfico [64](#)
    - Diálogo Imprimir [62](#)
    - Diálogo Inferência [59](#)
    - Diálogo Salvar [63](#)
    - Diálogo Viés [60](#)
    - Guia Gráfico de Bolhas [79](#)
    - Guia Gráfico de Floresta [77](#)
    - Guia Gráfico de Floresta Acumulativo [78](#)
    - Guia Gráfico de Funil [80](#)
    - Guia Gráfico de Galbraith [81](#)

Binário de Metanálise (*continuação*)  
  Guia Gráfico de L'Abbé [82](#)  
Binômio ajustado de meio-p  
  em proporções de amostras emparelhadas [110](#)  
Binômio exato  
  em proporções de amostras emparelhadas [110](#)  
Bonett-Price  
  em proporções de amostras emparelhadas [110](#)  
Bonferroni  
  no GLM [129](#)  
  no One-Way ANOVA  
  [120](#)  
boosting  
  em modelos lineares [141](#)  
boxplots  
  comparando níveis de fatores [92](#)  
  comparando variáveis [92](#)  
  na simulação [260](#)  
  no Explore [92](#)  
Brown-Li-Jeffreys  
  em Proporções de amostras independentes [113](#)

## C

C de Dunnett  
  no GLM [129](#)  
  no One-Way ANOVA  
  [120](#)  
camadas  
  em Tabulações Cruzadas [95](#)  
carregando gráficos  
  na Análise fatorial [183](#)  
categoria de referência  
  no GLM [126](#), [127](#)  
classificação  
  em Curva ROC [247](#)  
  na Análise de Característica de Operação do Receptor  
  [244](#)  
Clopper-Pearson (Exato)  
  em proporções de uma amostra [107](#)  
coeficiente alfa  
  na Análise de confiabilidade [227](#)  
coeficiente alpha  
  na Análise de confiabilidade [228](#)  
Coeficiente de concordância de Kendall (W)  
  Testes não paramétricos de amostras relacionadas [205](#)  
coeficiente de contingência  
  em Tabulações Cruzadas [95](#)  
coeficiente de correlação de classificação  
  em bivariate correlation [135](#)  
coeficiente de correlação de Spearman  
  em Tabulações Cruzadas [95](#)  
Coeficiente de correlação de Spearman  
  em bivariate correlation [135](#)  
coeficiente de correlação intraclasse (ICC)  
  na Análise de confiabilidade [228](#)  
coeficiente de correlação r  
  em bivariate correlation [135](#)  
  em Tabulações Cruzadas [95](#)  
coeficiente de dispersão (COD)  
  em Estatísticas de razão [236](#), [238](#)  
coeficiente de incerteza  
  em Tabulações Cruzadas [95](#)  
coeficiente de variação (COV)  
  em Estatísticas de razão [236](#), [238](#)  
coeficientes betas  
  Na Regressão Linear [150](#)  
coeficientes de regressão  
  Na Regressão Linear [150](#)  
coluna total  
  em relatórios [225](#)  
combinações  
  em modelos lineares [143](#)  
comparando grupos  
  em cubos OLAP [105](#)  
comparando variáveis  
  em cubos OLAP [105](#)  
Concordância entre avaliadores [228](#)  
confiabilidade de divisão de metade  
  na Análise de confiabilidade [228](#)  
confiabilidade dividida ao meio  
  na Análise de confiabilidade [227](#)  
Confiabilidade Spearman-Brown  
  na Análise de confiabilidade [228](#)  
conjuntos de múltiplas respostas  
  Livro de códigos [83](#)  
Construtor de simulação [251](#)  
contagem esperada  
  em Tabulações Cruzadas [97](#)  
contagem observada  
  em Tabulações Cruzadas [97](#)  
contrastes  
  no GLM [126](#), [127](#)  
  no One-Way ANOVA  
  [120](#)  
contrastes de desvio  
  no GLM [126](#), [127](#)  
contrastes de diferença  
  no GLM [126](#), [127](#)  
contrastes de Helmert  
  no GLM [126](#), [127](#)  
contrastes polinomiais  
  no GLM [126](#), [127](#)  
  no One-Way ANOVA  
  [120](#)  
contrastes repetidos  
  no GLM [126](#), [127](#)  
contrastes simples  
  no GLM [126](#), [127](#)  
controle de página  
  em relatórios de resumo de coluna [226](#)  
  relatórios de resumo em linha [224](#)  
convergência  
  na Análise de Cluster por K-Médias [196](#)  
  na Análise fatorial [182](#), [183](#)  
Correção de Yates para continuidade  
  em Tabulações Cruzadas [95](#)  
Correlação de Pearson  
  em bivariate correlation [135](#)  
  em Tabulações Cruzadas [95](#)  
correlações  
  em bivariate correlation [135](#)  
  em Correlações Parciais [137](#)  
  em Tabulações Cruzadas [95](#)  
  na simulação [257](#)  
  ordem zero [138](#)  
Correlações bivariadas

- Correlações bivariadas (*continuação*)
  - coeficientes de correlação [135](#)
  - Estatísticas [136](#)
  - intervalo de confiança [135](#)
  - intervalos de confiança [137](#)
  - nível de significância [135](#)
  - opções [136](#)
  - recursos adicionais do comando [137](#)
  - valores omissos [136](#)
- correlações de ordem zero
  - em Correlações Parciais [138](#)
- Correlações parciais
  - correlações de ordem zero [138](#)
  - estatísticas [138](#)
  - Na Regressão Linear [150](#)
  - opções [138](#)
  - recursos adicionais do comando [138](#)
  - valores omissos [138](#)
- Critério de informação de Akaike
  - em modelos lineares [142](#)
- critério de prevenção de super ajuste
  - em modelos lineares [142](#)
- critérios de informações
  - em modelos lineares [142](#)
- critérios de seleção
  - Na Regressão Linear [150](#)
- Cubos OLAP
  - estatísticas [104](#)
  - títulos [105](#)
- curtose
  - em cubos OLAP [104](#)
  - em Descritivos [89](#)
  - em Frequências [87](#)
  - em Médias [101](#)
  - em Resumos de Relatórios em Colunas [225](#)
  - em sumarizações de relatórios em linhas [223](#)
  - em Sumarizar [99](#)
  - no Explore [92](#)
- Curva ROC
  - estatísticas e gráficos [248](#)

## D

- d
  - em Tabulações Cruzadas [95](#)
- d de Somers
  - em Tabulações Cruzadas [95](#)
- decomposição hierárquica [125](#)
- Definir conjuntos de múltiplas respostas
  - categorias [219](#)
  - dicotomias [219](#)
  - nomes de conjunto [219](#)
  - rótulos do conjunto [219](#)
- dendrogramas
  - em Análise de Cluster Hierárquica [195](#)
- Descritivas
  - estatísticas [89](#)
  - ordem de exibição [89](#)
  - recursos adicionais do comando [90](#)
  - salvando pontuações z [89](#)
- designs para testes de heterocedasticidade
  - no GLM [134](#)
- desvio absoluto médio (AAD)
  - em Estatísticas de razão [236](#), [238](#)

- desvio padrão
  - em cubos OLAP [104](#)
  - em Descritivos [89](#)
  - em Estatísticas de razão [236](#), [238](#)
  - em Frequências [87](#)
  - em Médias [101](#)
  - em Resumos de Relatórios em Colunas [225](#)
  - em sumarizações de relatórios em linhas [223](#)
  - em Sumarizar [99](#)
  - no Explore [92](#)
- DfBeta
  - Na Regressão Linear [149](#)
- DfFit
  - Na Regressão Linear [149](#)
- Diálogo Análise
  - Binário de Metanálise [58](#)
  - Metanálise Contínua [41](#)
  - Tamanho do Efeito Binário de Metanálise [66](#)
  - Tamanho do Efeito Contínuo de Metanálise [49](#)
- Diálogo Aparar e Preencher
  - Binário de Metanálise [61](#)
  - Metanálise Contínua [44](#)
  - Tamanho do Efeito Binário de Metanálise [69](#)
  - Tamanho do Efeito Contínuo de Metanálise [52](#)
- Diálogo Contraste
  - Binário de Metanálise [60](#)
  - Metanálise Contínua [43](#)
  - Tamanho do Efeito Binário de Metanálise [68](#)
  - Tamanho do Efeito Contínuo de Metanálise [51](#)
- Diálogo Critérios
  - Binário de Metanálise [56](#)
  - Metanálise Contínua [40](#), [48](#)
  - Regressão de Metanálise [74](#)
  - Tamanho do Efeito Binário de Metanálise [65](#)
- Diálogo Gráfico
  - Binário de Metanálise [64](#)
  - Metanálise Contínua [47](#)
  - Regressão de Metanálise [76](#)
  - Tamanho do Efeito Binário de Metanálise [72](#)
  - Tamanho do Efeito Contínuo de Metanálise [55](#)
- Diálogo Imprimir
  - Binário de Metanálise [62](#)
  - Metanálise Contínua [46](#)
  - Regressão de Metanálise [75](#)
  - Tamanho do Efeito Binário de Metanálise [71](#)
  - Tamanho do Efeito Contínuo de Metanálise [54](#)
- Diálogo Inferência
  - Binário de Metanálise [59](#)
  - Metanálise Contínua [42](#)
  - Regressão de Metanálise [74](#)
  - Tamanho do Efeito Binário de Metanálise [67](#)
  - Tamanho do Efeito Contínuo de Metanálise [50](#)
- Diálogo Salvar
  - Binário de Metanálise [63](#)
  - Metanálise Contínua [46](#)
  - Regressão de Metanálise [76](#)
  - Tamanho do Efeito Binário de Metanálise [72](#)
  - Tamanho do Efeito Contínuo de Metanálise [54](#)
- Diálogo Viés
  - Binário de Metanálise [60](#)
  - Metanálise Contínua [44](#)
  - Tamanho do Efeito Binário de Metanálise [69](#)
  - Tamanho do Efeito Contínuo de Metanálise [52](#)
- dicionário

dicionário (*continuação*)  
   Livro de códigos [83](#)  
 diferença honestamente significativa de Tukey  
   no GLM [129](#)  
   no One-Way ANOVA [120](#)  
 diferença menos significativa  
   no GLM [129](#)  
   no One-Way ANOVA [120](#)  
 diferenças entre grupos  
   em cubos OLAP [105](#)  
 diferenças entre variáveis  
   em cubos OLAP [105](#)  
 diferencial relacionado ao preço (PRD)  
   em Estatísticas de razão [236](#), [238](#)  
 distância de bloco  
   em Distâncias [139](#)  
 distância de Chebychev  
   em Distâncias [139](#)  
 Distância de Cook  
   Na Regressão Linear [149](#)  
   no GLM [132](#)  
 Distância de Mahalanobis  
   na análise discriminante [179](#)  
   Na Regressão Linear [149](#)  
 distância de Manhattan  
   na Análise do vizinho mais próximo [171](#)  
 distância de Minkowski  
   em Distâncias [139](#)  
 distância do quarteirão  
   na Análise do vizinho mais próximo [171](#)  
 Distância euclidiana  
   em Distâncias [139](#)  
   na Análise do vizinho mais próximo [171](#)  
 distância euclidiana quadrática  
   em Distâncias [139](#)  
 distância qui-quadrado  
   em Distâncias [139](#)  
 Distâncias  
   calcular distâncias entre casos [139](#)  
   calcular distâncias entre variáveis [139](#)  
   estatísticas [139](#)  
   exemplo [139](#)  
   medidas de dissimilaridade [139](#)  
   medidas de similaridade [140](#)  
   medidas de transformação [139](#), [140](#)  
   recursos adicionais do comando [140](#)  
   valores de transformação [139](#), [140](#)  
 distâncias de vizinho mais próximo  
   na Análise do vizinho mais próximo [176](#)  
 diversas comparações de posteriori [120](#)  
 divisão  
   dividindo entre colunas do relatório [225](#)  
 Dois testes de amostras independentes  
   estatísticas [214](#)  
   opções [214](#)  
   tipos de teste [214](#)  
   valores omissos [214](#)

**E**

eliminação backward  
   Na Regressão Linear [148](#)  
 erro padrão  
   em Descritivos [89](#)  
   em Frequências [87](#)  
   na Análise de Característica de Operação do Receptor [246](#)  
   na Análise ROC [245](#)  
   na curva ROC [248](#)  
   no Explore [92](#)  
   no GLM [132](#)  
 erro padrão da média  
   em cubos OLAP [104](#)  
   em Médias [101](#)  
   em Sumarizar [99](#)  
 erro padrão de assimetria  
   em cubos OLAP [104](#)  
   em Médias [101](#)  
   em Sumarizar [99](#)  
 erro padrão de curtose  
   em cubos OLAP [104](#)  
   em Médias [101](#)  
   em Sumarizar [99](#)  
 escala  
   na Análise de confiabilidade [227](#)  
   no Ajuste de escala multidimensional [233](#)  
   no Kappa ponderado [231](#)  
 Escores do fator Anderson-Rubin [184](#)  
 escores dos fatores [184](#)  
 Estatística Brown-Forsythe  
   no One-Way ANOVA [122](#)  
 estatística de Cochran  
   em Tabulações Cruzadas [95](#)  
 estatística de Durbin-Watson  
   Na Regressão Linear [150](#)  
 Estatística de Mantel-Haenszel  
   em Tabulações Cruzadas [95](#)  
 Estatística de R  
   em Médias [101](#)  
 estatística de Welch  
   no One-Way ANOVA [122](#)  
 estatística descritiva  
   em Descritivos [89](#)  
   em Estatísticas de razão [236](#), [238](#)  
   em Frequências [87](#)  
   em Sumarizar [99](#)  
   na Análise de Cluster TwoStep [188](#)  
 estatística F  
   em modelos lineares [142](#)  
 estatística R  
   Na Regressão Linear [150](#)  
 estatísticas de proporções de coluna  
   em Tabulações Cruzadas [97](#)  
 Estatísticas de razão  
   estatísticas [236](#), [238](#)  
 estatísticas descritivas  
   no Explore [92](#)  
 Estimação de curva  
   modelos [166](#)  
 estimador de bponderação de Tukey  
   no Explore [92](#)  
 estimador de onda de Andrews  
   no Explore [92](#)  
 estimador M de Huber

- estimador M de Huber (*continuação*)
  - no Explore [92](#)
- Estimador M-estimador redescendente de Hampel
  - no Explore [92](#)
- Estimadores M
  - no Explore [92](#)
- Estimativa de curva
  - análise de variância [165](#)
  - incluindo constante [165](#)
  - previsão [166](#)
  - salvando intervalos de predição [166](#)
  - salvando resíduos [166](#)
  - salvando valores preditos [166](#)
- estimativas de Hodges-Lehman
  - Testes não paramétricos de amostras relacionadas [205](#)
- estimativas de parâmetro
  - em Regressão ordinal [153](#)
- estresse
  - no Ajuste de escala multidimensional [233](#)
- estresse de S
  - no Ajuste de escala multidimensional [233](#)
- estudo caso-controle
  - Teste-T de amostras em pares [116](#)
- estudo de pares
  - no teste T de amostras pareadas [116](#)
- eta
  - em Médias [101](#)
  - em Tabulações Cruzadas [95](#)
- eta quadrado
  - em Médias [101](#)
- Execuções de Wald-Wolfowitz
  - em Testes de duas amostras independentes [214](#)
- Explorar
  - estatísticas [92](#)
  - gráficos [92](#)
  - opções [93](#)
  - recursos adicionais do comando [93](#)
  - transformações de potência [93](#)
  - valores omissos [93](#)

## F

- F múltiplo de Ryan-Einot-Gabriel-Welsch
  - no GLM [129](#)
  - no One-Way ANOVA [120](#)
- fator de inflação de variância
  - Na Regressão Linear [150](#)
- fatoração alpha [182](#)
- fatoração de imagem [182](#)
- fatoração do eixo principal [182](#)
- fi
  - em Tabulações Cruzadas [95](#)
- first
  - em Médias [101](#)
- formatação
  - colunas em relatórios [223](#)
- forward stepwise
  - em modelos lineares [142](#)
- Frequências
  - estatísticas [87](#)
  - formatos [88](#)
  - gráficos [88](#)
  - ordem de exibição [88](#)

- Frequências (*continuação*)
  - suprimindo tabelas [88](#)
- frequências acumulativas
  - em Regressão ordinal [153](#)
- frequências de cluster
  - na Análise de Cluster TwoStep [188](#)
- Frequências de múltiplas respostas
  - valores omissos [219](#)
- frequências esperadas
  - em Regressão ordinal [153](#)
- frequências observadas
  - em Regressão ordinal [153](#)
- funções de distribuição cumulativas
  - na simulação [259](#)
- funções de densidade de probabilidade
  - na simulação [259](#)

## G

- gama
  - em Tabulações Cruzadas [95](#)
- gama de Goodman e Kruskal
  - em Tabulações Cruzadas [95](#)
- GLM
  - gráficos de perfil [127](#)
  - modelo [125](#)
  - salvando matrizes [132](#)
  - salvando variáveis [132](#)
  - soma dos quadrados [125](#)
  - testes post hoc [129](#)
- GLM Univariate
  - contrastes [126](#), [127](#)
- Gráfico de bolhas
  - Regressão de Metanálise [79](#)
- gráfico de dispersão
  - na simulação [260](#)
- gráfico de espaço da variável
  - na análise do vizinho mais próximo [174](#)
- Gráfico de floresta [77](#)
- Gráfico de floresta acumulativo [77](#)
- Gráfico de funil [77](#)
- Gráfico de Galbraith [77](#)
- Gráfico de L'Abbé [77](#)
- gráfico icicle
  - em Análise de Cluster Hierárquica [195](#)
- gráficos
  - em Curva ROC [247](#)
  - gráficos de probabilidade [239](#), [241](#)
  - na Análise de Característica de Operação do Receptor [244](#)
  - rótulos de caso [165](#)
- Gráficos
  - Metanálise [77](#)
- gráficos de barras
  - em Frequências [88](#)
- gráficos de dispersão
  - Na Regressão Linear [148](#)
- gráficos de dispersão versus nível
  - no Explore [92](#)
- gráficos de perfil
  - no GLM [127](#)
- gráficos de pizza
  - em Frequências [88](#)
- gráficos de probabilidade



gráficos de probabilidade (*continuação*)

P-P [239](#)

Q-Q [241](#)

gráficos de probabilidade normal

Na Regressão Linear [148](#)

no Explore [92](#)

P-P [239](#)

Q-Q [241](#)

Gráficos de probabilidade P-P [239](#)

Gráficos de probabilidade Q-Q [241](#)

gráficos de ramos e folhas

no Explore [92](#)

gráficos normais sem tendência

no Explore [92](#)

gráficos parciais

Na Regressão Linear [148](#)

gráficos tornado

na simulação [260](#)

GT2 de Hochberg

no GLM [129](#)

no One-Way ANOVA

[120](#)

Guia Gráfico de Bolhas

Binário de Metanálise [79](#)

Metanálise Contínua [79](#)

Tamanho do Efeito Binário de Metanálise [79](#)

Tamanho do Efeito Contínuo de Metanálise [79](#)

Guia Gráfico de Floresta

Binário de Metanálise [77](#)

Metanálise Contínua [77](#)

Tamanho do Efeito Binário de Metanálise [77](#)

Tamanho do Efeito Contínuo de Metanálise [77](#)

Guia Gráfico de Floresta Acumulativo

Binário de Metanálise [78](#)

Metanálise Contínua [78](#)

Tamanho do Efeito Binário de Metanálise [78](#)

Tamanho do Efeito Contínuo de Metanálise [78](#)

Guia Gráfico de Funil

Binário de Metanálise [80](#)

Metanálise Contínua [80](#)

Tamanho do Efeito Binário de Metanálise [80](#)

Tamanho do Efeito Contínuo de Metanálise [80](#)

Guia Gráfico de Galbraith

Binário de Metanálise [81](#)

Metanálise Contínua [81](#)

Tamanho do Efeito Binário de Metanálise [81](#)

Tamanho do Efeito Contínuo de Metanálise [81](#)

Guia Gráfico de L'Abbé

Binário de Metanálise [82](#)

Metanálise Contínua [82](#)

Tamanho do Efeito Binário de Metanálise [82](#)

Tamanho do Efeito Contínuo de Metanálise [82](#)

## H

H de Kruskal-Wallis

em Testes de duas amostras independentes [216](#)

Hauck-Anderson

em Proporções de amostras independentes [113](#)

histogramas

em Frequências [88](#)

Na Regressão Linear [148](#)

no Explore [92](#)

histórico de iteração

histórico de iteração (*continuação*)

em Regressão ordinal [153](#)

## I

ICC. Consulte coeficiente de correlação intraclasse [228](#)

importância de variável

na análise do vizinho mais próximo [176](#)

importância do preditor

modelos lineares [144](#)

informações de diagnóstico de colinearidade

Na Regressão Linear [150](#)

informações de diagnóstico entre casos

Na Regressão Linear [150](#)

intervalo de confiança

em bivariate correlation [135](#)

Intervalos de Clopper-Pearson

Testes não paramétricos de amostra única [199](#)

intervalos de confiança

em bivariate correlation [137](#)

em Teste T de Amostras Independentes [116](#)

em Teste T de uma amostra [118](#)

na Análise de Característica de Operação do Receptor [246](#)

na Análise ROC [245](#)

na curva ROC [248](#)

Na Regressão Linear [150](#)

no Explore [92](#)

no GLM [126](#)

no One-Way ANOVA [122](#)

no teste T de amostras pareadas [117](#)

salvando na regressão linear [149](#)

Intervalos de Jeffreys

Testes não paramétricos de amostra única [199](#)

intervalos de predição

salvando em Curva de Estimção [166](#)

salvando na regressão linear [149](#)

intervalos de razão de verossimilhança

Testes não paramétricos de amostra única [199](#)

iterações

na Análise de Cluster por K-Médias [196](#)

na Análise fatorial [182](#), [183](#)

## J

Jeffreys

em proporções de uma amostra [107](#)

## K

k e seleção de recursos

na análise do vizinho mais próximo [177](#)

kapa

em Tabulações Cruzadas [95](#)

Kappa de avaliador múltiplo de Fleiss [227](#), [228](#)

kappa de Cohen

em Tabulações Cruzadas [95](#)

Kappa ponderado

critérios [232](#)

estatísticas [232](#)

Estatísticas [231](#)

exemplo [231](#)

imprimir [232](#)

Kappa ponderado (*continuação*)  
  tabulação cruzada [232](#)  
Kappa ponderado de Cohen [231](#), [232](#)  
Kolmogorov-Smirnov Z  
  no Teste de Kolmogorov-Smirnov de uma amostra  
  [211](#)  
KR20  
  na Análise de confiabilidade [228](#)  
Kuder-Richardson 20 (KR20)  
  na Análise de confiabilidade [228](#)

## L

lambda  
  em Tabulações Cruzadas [95](#)  
lambda de Goodman e Kruskal  
  em Tabulações Cruzadas [95](#)  
Lambda de Wilks  
  na análise discriminante [179](#)  
limite inicial  
  na Análise de Cluster TwoStep [187](#)  
link  
  em Regressão ordinal [153](#)  
listando casos [98](#)  
Livro de códigos  
  estatísticas [86](#)  
  saída [84](#)  
Logit  
  em proporções de uma amostra [107](#)  
LSD de Fisher  
  no GLM [129](#)

## M

mapa de quadrante  
  na Análise do vizinho mais próximo [176](#)  
matriz de correlações  
  em Regressão ordinal [153](#)  
  na análise discriminante [179](#)  
  na Análise fatorial [181](#), [182](#)  
matriz de covariâncias  
  em Regressão ordinal [153](#)  
  na análise discriminante [179](#)  
  na Análise Discriminante [180](#)  
  Na Regressão Linear [150](#)  
  no GLM [132](#)  
matriz de padrão  
  na Análise fatorial [181](#)  
matriz de transformação  
  na Análise fatorial [181](#)  
máxima verossimilhança  
  na Análise fatorial [182](#)  
máximo  
  em cubos OLAP [104](#)  
  em Frequências [87](#)  
  em Sumarizar [99](#)  
  no Explore [92](#)  
máximo de ramificações  
  na Análise de Cluster TwoStep [187](#)  
maximum  
  comparando colunas de relatório [225](#)  
  em Descritivos [89](#)  
  em Estatísticas de razão [236](#), [238](#)

maximum (*continuação*)  
  em Médias [101](#)  
McNemar  
  em proporções de amostras emparelhadas [110](#)  
McNemar (continuidade corrigida)  
  em proporções de amostras emparelhadas [110](#)  
média  
  de diversas colunas de relatório [225](#)  
  em cubos OLAP [104](#)  
  em Descritivos [89](#)  
  em Estatísticas de razão [236](#), [238](#)  
  em Frequências [87](#)  
  em Médias [101](#)  
  em Resumos de Relatórios em Colunas [225](#)  
  em sumarizações de relatórios em linhas [223](#)  
  em Sumarizar [99](#)  
  no Explore [92](#)  
  no One-Way ANOVA [122](#)  
  subgrupo [101](#), [103](#)  
média aparada  
  no Explore [92](#)  
média geométrica  
  em cubos OLAP [104](#)  
  em Médias [101](#)  
  em Sumarizar [99](#)  
média harmônica  
  em cubos OLAP [104](#)  
  em Médias [101](#)  
  em Sumarizar [99](#)  
média ponderada  
  em Estatísticas de razão [236](#), [238](#)  
mediana  
  em cubos OLAP [104](#)  
  em Estatísticas de razão [236](#), [238](#)  
  em Frequências [87](#)  
  em Médias [101](#)  
  em Sumarizar [99](#)  
  no Explore [92](#)  
mediana agrupada  
  em cubos OLAP [104](#)  
  em Médias [101](#)  
  em Sumarizar [99](#)  
Médias  
  estatísticas [101](#)  
  opções [101](#)  
médias de grupo [101](#), [103](#)  
médias de subgrupo [101](#), [103](#)  
medida de diferença de tamanho  
  em Distâncias [139](#)  
medida de diferença padrão  
  em Distâncias [139](#)  
medida de dissimilaridade de Lance e Williams  
  em Distâncias [139](#)  
medida de distância fi-quadrado  
  em Distâncias [139](#)  
medidas de dispersão  
  em Descritivos [89](#)  
  em Estatísticas de razão [236](#), [238](#)  
  em Frequências [87](#)  
  no Explore [92](#)  
medidas de distância  
  em Análise de Cluster Hierárquica [194](#)  
  em Distâncias [139](#)  
  na Análise do vizinho mais próximo [171](#)



medidas de distribuição  
     em Descritivos [89](#)  
     em Frequências [87](#)  
 medidas de similaridade  
     em Análise de Cluster Hierárquica [194](#)  
     em Distâncias [140](#)  
 medidas de tendência central  
     em Estatísticas de razão [236](#), [238](#)  
     em Frequências [87](#)  
     no Explore [92](#)  
 melhores subconjuntos  
     em modelos lineares [142](#)  
 Meta-análise  
     regressão [38](#)  
     resultados binários [38](#)  
     resultados contínuos [38](#)  
 Metanálise  
     Gráficos [77](#)  
 Metanálise Contínua  
     Diálogo Análise [41](#)  
     Diálogo Aparar e Preencher [44](#)  
     Diálogo Contraste [43](#)  
     Diálogo Critérios [40](#), [48](#)  
     Diálogo Gráfico [47](#)  
     Diálogo Imprimir [46](#)  
     Diálogo Inferência [42](#)  
     Diálogo Salvar [46](#)  
     Diálogo Viés [44](#)  
     Guia Gráfico de Bolhas [79](#)  
     Guia Gráfico de Floresta [77](#)  
     Guia Gráfico de Floresta Acumulativo [78](#)  
     Guia Gráfico de Funil [80](#)  
     Guia Gráfico de Galbraith [81](#)  
     Guia Gráfico de L'Abbé [82](#)  
 mínimo  
     comparando colunas de relatório [225](#)  
     em cubos OLAP [104](#)  
     em Descritivos [89](#)  
     em Estatísticas de razão [236](#), [238](#)  
     em Frequências [87](#)  
     em Sumarizar [99](#)  
     no Explore [92](#)  
 Mínimo  
     em Médias [101](#)  
 mínimos quadrados generalizados  
     na Análise fatorial [182](#)  
 mínimos quadrados não ponderados  
     na Análise fatorial [182](#)  
 mínimos quadrados ponderados  
     Na Regressão Linear [147](#)  
 modelagem espacial [265](#)  
 modelagem geoespacial [265](#)–[276](#)  
 modelo composto  
     na Estimativa de curva [166](#)  
 modelo cúbico  
     na Estimativa de curva [166](#)  
 modelo de crescimento  
     na Estimativa de curva [166](#)  
 modelo de escala  
     em Regressão Ordinal [155](#)  
 modelo de Guttman  
     na Análise de confiabilidade [228](#)  
 Modelo de Guttman  
     na Análise de confiabilidade [227](#)  
 modelo de localização  
     em Regressão Ordinal [154](#)  
 modelo de potência  
     na Estimativa de curva [166](#)  
 modelo de regressão auxiliar  
     no GLM [134](#)  
 modelo estritamente paralelo  
     na Análise de confiabilidade [227](#), [228](#)  
 modelo exponencial  
     na Estimativa de curva [166](#)  
 modelo inverso  
     na Estimativa de curva [166](#)  
 modelo linear  
     na Estimativa de curva [166](#)  
 modelo logarítmico  
     na Estimativa de curva [166](#)  
 modelo logístico  
     na Estimativa de curva [166](#)  
 modelo paralelo  
     na Análise de confiabilidade [227](#), [228](#)  
 modelo quadrático  
     na Estimativa de curva [166](#)  
 modelo S  
     na Estimativa de curva [166](#)  
 modelos customizados  
     no GLM [125](#)  
 modelos fatoriais completos  
     no GLM [125](#)  
 modelos lineares  
     coeficientes [145](#)  
     combinações [143](#)  
     critério de informação [143](#)  
     Estatística de R-quadrado [143](#)  
     importância do preditor [144](#)  
     médias estimadas [146](#)  
     nível de confiança [141](#)  
     objetivos [141](#)  
     opções de modelo [143](#)  
     predito por observado [144](#)  
     preparação automática de dados [144](#)  
     preparação de dado automático [141](#)  
     regras de combinação [143](#)  
     replicando resultados [143](#)  
     resíduos [144](#)  
     resumo do modelo [143](#)  
     seleção de modelo [142](#)  
     sumarização de construção de modelo [146](#)  
     Tabela de ANOVA [145](#)  
     valores discrepantes [145](#)  
 modo  
     em Frequências [87](#)  
 múltiplas comparações  
     no One-Way ANOVA  
     [120](#)  
 Múltiplas respostas  
     recursos adicionais do comando [222](#)  
 multiplicação  
     multiplicando entre colunas do relatório [225](#)  
**N**  
 Newcombe  
     em proporções de amostras emparelhadas [110](#)  
     em Proporções de amostras independentes [113](#)

Newcombe (continuidade corrigida)  
em Proporções de amostras independentes [113](#)  
Newman-Keuls  
no GLM [129](#)  
Nível de significância assintótico [228](#)  
Normalização de Kaiser  
na Análise fatorial [183](#)  
numeração de página  
em relatórios de resumo de coluna [226](#)  
relatórios de resumo em linha [224](#)  
número de casos  
em cubos OLAP [104](#)  
em Médias [101](#)  
em Sumarizar [99](#)

## O

O teste exato de Fisher  
em Tabulações Cruzadas [95](#)

## P

padronização  
na Análise de Cluster TwoStep [187](#)  
peers  
na análise do vizinho mais próximo [176](#)  
percentis  
em Frequências [87](#)  
na simulação [260](#)  
no Explore [92](#)  
Percentis [90](#), [91](#)  
PLUM  
em Regressão ordinal [152](#)  
Pontuação [108](#)  
Pontuação (Continuidade corrigida) [108](#)  
Pontuação de Wilson  
em proporções de uma amostra [107](#)  
Pontuações do fator Bartlett [184](#)  
pontuações z  
em Descritivos [89](#)  
salvando como variáveis [89](#)  
porcentagens  
em Tabulações Cruzadas [97](#)  
porcentagens de coluna  
em Tabulações Cruzadas [97](#)  
porcentagens de linha  
em Tabulações Cruzadas [97](#)  
porcentagens totais  
em Tabulações Cruzadas [97](#)  
preparação automática de dados  
em modelos lineares [144](#)  
previsão  
na Curva de estimação [166](#)  
Primeiro  
em cubos OLAP [104](#)  
em Sumarizar [99](#)  
profundidade da árvore  
na Análise de Cluster TwoStep [187](#)  
Proporções de amostras independentes [111](#), [113](#)  
Proporções de amostras pareadas [109](#)  
Proporções de uma amostra [106](#), [107](#), [110](#)  
Proximidades  
em Análise de Cluster Hierárquica [193](#)

## Q

Q de Cochran  
em Testes para várias amostras relacionadas [217](#)  
Qualidade do ajuste  
em Regressão ordinal [153](#)  
quartis  
em Frequências [87](#)  
qui-quadrado  
associação linear por linear [95](#)  
Correção de Yates para continuidade [95](#)  
em Tabulações Cruzadas [95](#)  
estatísticas [209](#)  
intervalo esperado [208](#)  
O teste exato de Fisher [95](#)  
opções [209](#)  
para a independência [95](#)  
Pearson [95](#)  
razão de verossimilhança [95](#)  
teste de uma amostra [208](#)  
valores esperados [208](#)  
valores omissos [209](#)  
Qui-quadrado de Pearson  
em Regressão ordinal [153](#)  
em Tabulações Cruzadas [95](#)  
qui-quadrado de razão de verossimilhança  
em Regressão ordinal [153](#)  
em Tabulações Cruzadas [95](#)

## R

R<sup>2</sup>  
em Médias [101](#)  
mudança de R<sup>2</sup> [150](#)  
Na Regressão Linear [150](#)  
R<sup>2</sup> ajustado  
Na Regressão Linear [150](#)  
R múltiplo  
Na Regressão Linear [150](#)  
R-E-G-W F  
no GLM [129](#)  
no One-Way ANOVA  
[120](#)  
R-E-G-W Q  
no GLM [129](#)  
no One-Way ANOVA  
[120](#)  
R-quadrado  
em modelos lineares [143](#)  
R-quadrado ajustado  
em modelos lineares [142](#)  
R<sup>2</sup> de Cox e Snell  
em Regressão ordinal [153](#)  
R<sup>2</sup> de McFadden  
em Regressão ordinal [153](#)  
R<sup>2</sup> de Nagelkerke  
em Regressão ordinal [153](#)  
razão de covariância  
Na Regressão Linear [149](#)  
regras de combinação  
em modelos lineares [143](#)  
regressão  
gráficos [148](#)  
Meta-análise [38](#)

regressão (*continuação*)  
     Regressão linear [147](#)  
     Regressão múltipla [147](#)  
 Regressão de Metanálise  
     Diálogo Critérios [74](#)  
     Diálogo Gráfico [76](#)  
     Diálogo Imprimir [75](#)  
     Diálogo Inferência [74](#)  
     Diálogo Salvar [76](#)  
     Gráfico de bolhas [79](#)  
 Regressão de rede elástica linear  
     opções [156](#)  
     partição [156](#)  
     variáveis [156](#)  
 Regressão de Ridge linear  
     opções [163](#)  
     partição [162](#)  
     variáveis [162](#)  
 Regressão linear  
     bloqueios [147](#)  
     exportando informações de modelo [149](#)  
     gráficos [148](#)  
     métodos de seleção de variáveis [148](#)  
     pesos [147](#)  
     recursos adicionais do comando [152](#)  
     resíduos [149](#)  
     salvando novas variáveis [149](#)  
     variável de seleção [148](#)  
 Regressão Linear  
     estatísticas [150](#)  
     métodos de seleção de variáveis [151](#)  
     valores omissos [151](#)  
 Regressão linear lasso  
     opções [160](#)  
     partição [159](#)  
     variáveis [159](#)  
 Regressão múltipla  
     Na Regressão Linear [147](#)  
 Regressão ordinal  
     estatísticas [152](#)  
     link [153](#)  
     modelo de escala [155](#)  
     modelo de localização [154](#)  
     options [153](#)  
     recursos adicionais do comando [155](#)  
 Regressão por quadrados mínimos parciais  
     modelo [169](#)  
     variáveis de exportação [169](#)  
 relatórios  
     colunas totais [225](#)  
     comparando colunas [225](#)  
     dividindo valores de coluna [225](#)  
     multiplicando valores de coluna [225](#)  
     relatórios de resumo de coluna [224](#)  
     relatórios sumarização em linha [222](#)  
     totais compostos [225](#)  
 relatórios de resumo de coluna [224](#)  
 resíduos  
     em Tabulações Cruzadas [97](#)  
     salvando em Curva de Estimção [166](#)  
     salvando na regressão linear [149](#)  
 Resíduos de Pearson  
     em Regressão ordinal [153](#)  
 Resíduos estudentizados  
     Resíduos estudentizados (*continuação*)  
         Na Regressão Linear [149](#)  
 resíduos excluídos  
     Na Regressão Linear [149](#)  
     no GLM [132](#)  
 resíduos não padronizados  
     no GLM [132](#)  
 resíduos padronizados  
     Na Regressão Linear [149](#)  
     no GLM [132](#)  
 resultados binários  
     dados brutos de tamanho do efeito [38](#)  
     dados de tamanho do efeito pré-calculado [38](#)  
 resultados contínuos  
     dados brutos de tamanho do efeito [38](#)  
     dados de tamanho do efeito pré-calculado [38](#)  
     Meta-análise [38](#)  
 Resumir  
     estatísticas [99](#)  
     opções [99](#)  
 Resumos de relatórios em colunas  
     colunas totais [225](#)  
     controle de página [226](#)  
     formato de coluna [223](#)  
     layout da página [224](#)  
     numeração de página [226](#)  
     subtotais [226](#)  
     total geral [226](#)  
     valores omissos [226](#)  
 Resumos de relatórios em Colunas  
     recursos adicionais do comando [226](#)  
 Resumos de relatórios em linhas  
     formato de coluna [223](#)  
     layout da página [224](#)  
     rodapés [224](#)  
     títulos [224](#)  
     variáveis em títulos [224](#)  
 Resumos de relatórios em Linhas  
     controle de página [223](#)  
     espaçamento de quebra [223](#)  
     numeração de página [224](#)  
     recursos adicionais do comando [226](#)  
     valores omissos [224](#)  
 risco  
     em Tabulações Cruzadas [95](#)  
 risco relativo  
     em Tabulações Cruzadas [95](#)  
 rô  
     em bivariate correlation [135](#)  
     em Tabulações Cruzadas [95](#)  
 rotação equamax  
     na Análise fatorial [183](#)  
 rotação oblimin direta  
     na Análise fatorial [183](#)  
 rotação quartimax  
     na Análise fatorial [183](#)  
 rotação varimax  
     na Análise fatorial [183](#)  
  
**S**  
 seleção de variável  
     na Análise do vizinho mais próximo [177](#)  
 seleção forward

- seleção forward (*continuação*)
    - na análise do vizinho mais próximo [172](#)
    - Na Regressão Linear [148](#)
  - Seleção k
    - na Análise do vizinho mais próximo [177](#)
  - seleção stepwise
    - Na Regressão Linear [148](#)
  - simulação
    - ajuste de distribuição [254](#)
    - amostragem de cauda [258](#)
    - análise de sensibilidade [257](#)
    - análise what-if [257](#)
    - box plots [260](#)
    - Construtor de simulação [251](#)
    - correlações entre entradas [257](#)
    - criando novas entradas [253](#)
    - criando um plano de simulação [249](#), [250](#)
    - critérios de parada [258](#)
    - customizando o ajuste de distribuição [256](#)
    - editor de equação [252](#)
    - especificação de modelo [251](#)
    - executando um plano de simulação [250](#), [262](#)
    - formatos de exibição para destinos e entradas [260](#)
    - função de densidade de probabilidade [259](#)
    - função de distribuição cumulativa [259](#)
    - gráficos de dispersão [260](#)
    - gráficos interativos [264](#)
    - gráficos tornado [260](#)
    - modelos suportados [251](#)
    - opções de gráfico [265](#)
    - percentis de distribuição de resposta [260](#)
    - reajustando distribuições para novos dados [262](#)
    - resultados do ajuste de distribuição [256](#)
    - saída [259](#), [260](#)
    - salvar dados simulados [261](#)
    - salvar plano de simulação [261](#)
  - simulação de Monte Carlo [248](#)
  - soma
    - em cubos OLAP [104](#)
    - em Descritivos [89](#)
    - em Frequências [87](#)
    - em Médias [101](#)
    - em Sumarizar [99](#)
  - soma dos quadrados
    - no GLM [125](#)
  - Student-Newman-Keuls
    - no GLM [129](#)
    - no One-Way ANOVA [120](#)
  - subtotais
    - em relatórios de resumo de coluna [226](#)
  - sumarização de erro
    - na análise do vizinho mais próximo [177](#)
  - Sumarizações de Relatórios em Linhas
    - colunas de dados [222](#)
    - colunas de quebra [222](#)
    - seqüências de classificação [222](#)
- T**
- T 2 de Hotelling
    - na Análise de confiabilidade [227](#), [228](#)
  - T2 de Tamhane
    - no GLM [129](#)
  - T2 de Tamhane (*continuação*)
    - no One-Way ANOVA [120](#)
  - T3 de Dunnett
    - no GLM [129](#)
    - no One-Way ANOVA [120](#)
  - tabela de classificação
    - na análise do vizinho mais próximo [177](#)
  - Tabelas cruzadas de múltiplas respostas
    - definindo intervalos de valores [221](#)
  - tabelas de contingência [94](#)
  - tabelas de frequência
    - em Frequências [86](#)
    - no Explore [92](#)
  - tabulação cruzada
    - em Tabulações Cruzadas [94](#)
    - múltiplas respostas [220](#)
  - Tabulações cruzadas
    - camadas [95](#)
    - estatísticas [95](#)
    - exibição da célula [97](#)
    - formatos [98](#)
    - gráficos de barras em cluster [95](#)
    - suprimindo tabelas [94](#)
    - variáveis de controle [95](#)
  - Tabulações cruzadas de múltiplas respostas
    - correspondendo variáveis em conjuntos de respostas [221](#)
    - percentagens com base em casos [221](#)
    - percentagens com base em respostas [221](#)
    - percentagens de células [221](#)
    - valores omissos [221](#)
  - tamanho do efeito
    - em Teste T de Amostras Independentes [114](#)
    - em teste T de amostras pareadas [116](#)
  - Tamanho do Efeito Binário de Metanálise
    - Diálogo Análise [66](#)
    - Diálogo Aparar e Preencher [69](#)
    - Diálogo Contraste [68](#)
    - Diálogo Critérios [65](#)
    - Diálogo Gráfico [72](#)
    - Diálogo Imprimir [71](#)
    - Diálogo Inferência [67](#)
    - Diálogo Salvar [72](#)
    - Diálogo Viés [69](#)
    - Guia Gráfico de Bolhas [79](#)
    - Guia Gráfico de Floresta [77](#)
    - Guia Gráfico de Floresta Acumulativo [78](#)
    - Guia Gráfico de Funil [80](#)
    - Guia Gráfico de Galbraith [81](#)
    - Guia Gráfico de L'Abbé [82](#)
  - Tamanho do Efeito Contínuo de Metanálise
    - Diálogo Análise [49](#)
    - Diálogo Aparar e Preencher [52](#)
    - Diálogo Contraste [51](#)
    - Diálogo Gráfico [55](#)
    - Diálogo Imprimir [54](#)
    - Diálogo Inferência [50](#)
    - Diálogo Salvar [54](#)
    - Diálogo Viés [52](#)
    - Guia Gráfico de Bolhas [79](#)
    - Guia Gráfico de Floresta [77](#)
    - Guia Gráfico de Floresta Acumulativo [78](#)

- Tamanho do Efeito Contínuo de Metanálise (*continuação*)  
 Guia Gráfico de Funil [80](#)  
 Guia Gráfico de Galbraith [81](#)  
 Guia Gráfico de L'Abbé [82](#)
- tau de Goodman e Kruskal  
 em Tabulações Cruzadas [95](#)
- tau de Kruskal  
 em Tabulações Cruzadas [95](#)
- tau-b  
 em Tabulações Cruzadas [95](#)
- Tau-b de Kendall  
 em bivariate correlation [135](#)  
 em Tabulações Cruzadas [95](#)
- tau-c  
 em Tabulações Cruzadas [95](#)
- Tau-c de Kendall  
 em Tabulações Cruzadas [95](#)
- termos de construção [125](#), [155](#)
- termos de interação [125](#), [155](#)
- teste B de Tukey  
 no GLM [129](#)  
 no One-Way ANOVA [120](#)
- teste binomial  
 Testes não paramétricos de amostra única [199](#)
- Teste binomial  
 dicotomias [209](#)
- Teste de aditividade de Tukey  
 na Análise de confiabilidade [227](#), [228](#)
- teste de amplitude múltipla de Duncan  
 no GLM [129](#)  
 no One-Way ANOVA [120](#)
- Teste de binômio  
 estatísticas [210](#)  
 opções [210](#)  
 recursos adicionais do comando [210](#)  
 valores omissos [210](#)
- teste de comparação entre pares de Gabriel  
 no GLM [129](#)  
 no One-Way ANOVA [120](#)
- Teste de comparações pairwise de Games-Howell  
 no GLM [129](#)  
 no One-Way ANOVA [120](#)
- teste de esfericidade de Bartlett  
 na Análise fatorial [182](#)
- Teste de Friedman  
 em Testes para várias amostras relacionadas [217](#)  
 Testes não paramétricos de amostras relacionadas [205](#)
- teste de homogeneidade marginal  
 em Testes de Duas Amostras Relacionadas [215](#)  
 Testes não paramétricos de amostras relacionadas [205](#)
- teste de Kolmogorov-Smirnov  
 Testes não paramétricos de amostra única [199](#)  
 Testes não paramétricos de uma amostra [200](#)
- Teste de Kolmogorov-Smirnov  
 Teste de Lilliefors [200](#)
- Teste de Kolmogorov-Smirnov de uma amostra  
 amostra  
 distribuição de teste [211](#)  
 estatísticas [212](#)  
 Estatísticas [213](#)  
 opções [212](#), [213](#)  
 recursos adicionais do comando [213](#)
- Teste de Kolmogorov-Smirnov de uma amostra (*continuação*)  
 Teste de Lilliefors [211](#), [212](#)  
 valores omissos [212](#), [213](#)
- Teste de Levene  
 no Explore [92](#)  
 no One-Way ANOVA [122](#)
- teste de Lilliefors  
 no Explore [92](#)
- Teste de Lilliefors [200](#), [211](#), [212](#)
- teste de linhas paralelas  
 em Regressão ordinal [153](#)
- teste de McNemar  
 em Tabulações Cruzadas [95](#)  
 em Testes de Duas Amostras Relacionadas [215](#)
- Teste de McNemar  
 Testes não paramétricos de amostras relacionadas [205](#), [206](#)
- teste de mediana  
 em Testes de duas amostras independentes [216](#)
- Teste de reação extrema de Moses  
 em Testes de duas amostras independentes [214](#)
- teste de Scheffé  
 no GLM [129](#)  
 no One-Way ANOVA [120](#)
- teste de sequência  
 Testes não paramétricos de amostra única [199](#), [200](#)
- Teste de sequência  
 recursos adicionais do comando [211](#)
- Teste de Sequência  
 estatísticas [211](#)  
 opções [211](#)  
 pontos de corte [210](#), [211](#)  
 valores omissos [211](#)
- teste de Shapiro-Wilk  
 no Explore [92](#)
- teste de sinal  
 em Testes de Duas Amostras Relacionadas [215](#)  
 Testes não paramétricos de amostras relacionadas [205](#)
- Teste dos postos sinalizados de Wilcoxon  
 em Testes de Duas Amostras Relacionadas [215](#)  
 Testes não paramétricos de amostra única [199](#)  
 Testes não paramétricos de amostras relacionadas [205](#)
- Teste M de Box  
 na análise discriminante [179](#)
- teste Q de Cochran  
 Testes não paramétricos de amostras relacionadas [205](#), [206](#)
- teste qui-quadrado  
 Testes não paramétricos de amostra única [199](#)  
 Testes não paramétricos de uma amostra [200](#)
- Teste t  
 em Teste T de Amostras Independentes [114](#)  
 em Teste T de uma amostra [117](#)  
 no teste T de amostras pareadas [116](#)
- Teste T de amostras independentes  
 Definindo grupos [115](#)  
 intervalos de confiança [116](#)  
 opções [116](#)  
 valores omissos [116](#)  
 variáveis de agrupamento [115](#)  
 variáveis de sequência de caracteres [115](#)
- Teste t de amostras pareadas [116](#)

- teste t de duas amostras
  - em Teste T de Amostras Independentes [114](#)
- teste T de Dunnett
  - no GLM [129](#)
  - no One-Way ANOVA [120](#)
- teste T de Sidak
  - no GLM [129](#)
  - no One-Way ANOVA [120](#)
- Teste t de Student [114](#)
- teste T de Waller-Duncan
  - no GLM [129](#)
  - no One-Way ANOVA [120](#)
- teste t dependente
  - no teste T de amostras pareadas [116](#)
- Teste-T de amostras em pares
  - opções [117](#)
  - selecionando variáveis pareadas [116](#)
  - valores omissos [117](#)
- Teste-T de uma amostra
  - intervalos de confiança [118](#)
  - opções [118](#)
  - recursos adicionais do comando [117–119](#)
  - valores omissos [118](#)
- Testes de duas amostras independentes
  - definindo grupos [214](#)
  - recursos adicionais do comando [215](#)
  - variáveis de agrupamento [214](#)
- Testes de duas amostras relacionadas
  - estatísticas [215](#)
  - opções [215](#)
  - recursos adicionais do comando [216](#)
  - tipos de teste [215](#)
  - valores omissos [215](#)
- testes de homogeneidade de variância
  - no One-Way ANOVA [122](#)
- testes de independência
  - qui-quadrado [95](#)
- testes de linearidade
  - em Médias [101](#)
- testes de normalidade
  - no Explore [92](#)
- testes não paramétricos
  - qui-quadrado [208](#)
  - Teste de Kolmogorov-Smirnov de uma amostra [211](#)
  - Teste de Sequência [210](#)
  - Testes de duas amostras independentes [213](#)
  - Testes de duas amostras relacionadas [215](#)
  - Testes para diversas amostras independentes [216](#)
  - Testes para várias amostras relacionadas [217](#)
- Testes não paramétricos de amostra única
  - teste binomial [199](#)
  - teste de sequência [200](#)
- Testes Não Paramétricos de Amostras Independentes
  - Guia Campos [202](#)
- Testes não paramétricos de amostras relacionadas
  - campos [205](#)
  - Teste de McNemar [206](#)
  - teste Q de Cochran [206](#)
- Testes não paramétricos de uma amostra
  - campos [198](#)

- Testes não paramétricos de uma amostra (*continuação*)
  - teste de Kolmogorov-Smirnov [200](#)
  - teste qui-quadrado [200](#)
- Testes para diversas amostras independentes
  - definindo intervalo [217](#)
  - estatísticas [217](#)
  - opções [217](#)
  - recursos adicionais do comando [217](#)
  - tipos de teste [216](#)
  - valores omissos [217](#)
  - variáveis de agrupamento [217](#)
- Testes para várias amostras relacionadas
  - estatísticas [218](#)
  - recursos adicionais do comando [218](#)
  - tipos de teste [217](#)
- títulos
  - em cubos OLAP [105](#)
- tolerância
  - Na Regressão Linear [150](#)
- Tolerância
  - Regressão Linear [151](#)
- totais gerais
  - em relatórios de resumo de coluna [226](#)
- tratamento de ruído
  - na Análise de Cluster TwoStep [187](#)

**U**

- U de Mann-Whitney
  - em Testes de duas amostras independentes [214](#)

**V**

- V
  - em Tabulações Cruzadas [95](#)
- V de Cramér
  - em Tabulações Cruzadas [95](#)
- V de Rao
  - na análise discriminante [179](#)
- valores de leverage
  - Na Regressão Linear [149](#)
  - no GLM [132](#)
- valores discrepantes
  - na Análise de Cluster TwoStep [187](#)
  - Na Regressão Linear [148](#)
  - no Explore [92](#)
- valores extremos
  - no Explore [92](#)
- valores omissos
  - em bivariate correlation [136](#)
  - em Correlações Parciais [138](#)
  - em Frequências de múltiplas respostas [219](#)
  - em proporções de amostras emparelhadas [111](#)
  - em Proporções de amostras independentes [114](#)
  - em proporções de uma amostra [108](#)
  - em relatórios de resumo de coluna [226](#)
  - em Resumos de Relatórios em Linhas [224](#)
  - em Tabulações cruzadas de múltiplas respostas [221](#)
  - em Teste binomial [210](#)
  - em Teste de sequência [211](#)
  - em Teste qui-quadrado [209](#)
  - em Teste T de Amostras Independentes [116](#)
  - em Teste T de uma amostra [118](#)



- valores omissos (*continuação*)
  - em Testes de duas amostras independentes [214](#)
  - em Testes de Duas Amostras Relacionadas [215](#)
  - em Testes para várias amostras independentes [217](#)
  - na Análise de Característica de Operação do Receptor [246](#)
  - na análise do vizinho mais próximo [174](#)
  - na Análise fatorial [184](#)
  - na Análise ROC [245](#)
  - na curva ROC [248](#)
  - Na Regressão Linear [151](#)
  - no Explore [93](#)
  - no One-Way ANOVA [122](#)
  - no teste de Kolmogorov-Smirnov de uma amostra [212](#)
  - no Teste de Kolmogorov-Smirnov de uma amostra [213](#)
  - no teste T de amostras pareadas [117](#)
- Valores omissos
  - em Percentis [91](#)
- valores padronizados
  - em Descritivos [89](#)
- valores preditos
  - salvando em Curva de Estimção [166](#)
  - salvando na regressão linear [149](#)
- valores preditos ponderados
  - no GLM [132](#)
- variação
  - em cubos OLAP [104](#)
  - em Descritivos [89](#)
  - em Frequências [87](#)
  - em Médias [101](#)
  - em Resumos de Relatórios em Colunas [225](#)
  - em sumarizações de relatórios em linhas [223](#)
  - em Sumarizar [99](#)
  - no Explore [92](#)
- variáveis de controle
  - em Tabulações Cruzadas [95](#)
- variável de seleção
  - Na Regressão Linear [148](#)
- visualização
  - modelos de armazenamento em cluster [189](#)
- visualização do modelo
  - na análise do vizinho mais próximo [174](#)
- visualizador de cluster
  - classificação de exibição de cluster [191](#)
  - classificação de exibição de recursos [190](#)
  - comparação de clusters [192](#)
  - distribuição de células [192](#)
  - exibição de conteúdo da célula [191](#)
  - filtrando registros [193](#)
  - importância do preditor [191](#)
  - inverter clusters e recursos [190](#)
  - ordenar clusters [191](#)
  - ordenar conteúdo da célula [191](#)
  - ordenar variáveis [190](#)
  - resumo do modelo [189](#)
  - sobre modelos de cluster [188](#)
  - tamanho de clusters [191](#)
  - transpor clusters e recursos [190](#)
  - utilização [192](#)
  - visão geral [189](#)
  - visualização básica [191](#)
  - visualização de centros de cluster [190](#)
  - visualização de clusters [190](#)
  - visualização de comparação de cluster [192](#)

- visualizador de cluster (*continuação*)
  - visualização de distribuição de células [192](#)
  - visualização de importância do preditor de cluster [191](#)
  - visualização de sumarização [189](#)
  - visualização de tamanhos de cluster [191](#)

## W

- W de Kendall
  - em Testes para várias amostras relacionadas [217](#)
- Wald
  - em proporções de amostras emparelhadas [110](#)
  - em Proporções de amostras independentes [113](#)
  - em proporções de uma amostra [107](#)
- Wald (continuidade corrigida)
  - em proporções de amostras emparelhadas [110](#)
  - em Proporções de amostras independentes [113](#)
  - em proporções de uma amostra [107](#)
- Wald (Continuidade corrigida)
  - em proporções de amostras emparelhadas [110](#)
  - em Proporções de amostras independentes [113](#)
- Wald H0
  - em Proporções de amostras independentes [113](#)
- Wald H0 (continuidade corrigida)
  - em Proporções de amostras independentes [113](#)
- Wilson Score (continuidade corrigida)
  - em proporções de uma amostra [107](#)

## Z

- Z de Kolmogorov-Smirnov
  - em Testes de duas amostras independentes [214](#)







