

*IBM SPSS Advanced Statistics 31*



**Observação**

Antes de usar estas informações e o produto suportado por elas, leia as informações nos “Avisos” na [página 141](#).

**Informações sobre o produto**

Essa edição se aplica à versão 31, liberação 0, modificação 1 de IBM® SPSS Estatísticas e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

© Copyright International Business Machines Corporation .

<b>Capítulo 1. Estatísticas avançadas.....</b>	<b>1</b>
Introdução às Estatísticas Avançadas.....	1
Análise do GLM Multivariado.....	1
Modelo GLM Multivariate.....	3
Contrastes do GLM Multivariate.....	4
Gráficos de Perfil do GLM Multivariate.....	5
Comparações Post Hoc do GLM Multivariate.....	5
Médias marginais estimadas de GLM.....	6
Salvamento de GLM.....	6
Opções do GLM Multivariado.....	7
Recursos Adicionais de Comando do GLM.....	9
GLM Repeated Measures.....	9
Medidas Repetidas do GLM - Definir Fatores.....	11
Modelo de Medidas Repetidas do GLM.....	12
Contrastes do GLM Repeated Measures.....	13
Contrastes de Perfil do GLM Repeated Measures.....	14
Comparações Post Hoc do GLM Repeated Measures.....	14
Médias marginais estimadas de GLM.....	15
Salvamento de Medidas Repetidas do GLM.....	16
Opções de Medidas Repetidas do GLM.....	16
Recursos Adicionais de Comando do GLM.....	17
Análise de Componentes de Variância.....	18
Modelos de Componentes de Variância.....	19
Opções de Componentes de Variância.....	20
Componentes de Variância Salvos no Novo Arquivo.....	21
Recursos Adicionais do Comando VARCOMP.....	21
Modelos mistos lineares.....	21
Modelos lineares mistos: Assuntos e Repetidos.....	22
Efeitos Fixos de Modelos Lineares Mistos.....	24
Efeitos Aleatórios de Modelos Lineares Mistos.....	25
Estimação de Modelos Lineares.....	26
Estatísticas de Modelos Lineares Mistos.....	27
Médias de EM de Modelos Lineares Mistos.....	28
Salvamento de Modelos Lineares Mistos.....	28
Modelos Lineares Combinados-Exportação.....	28
Recursos Adicionais do Comando MIXED.....	28
Modelos lineares generalizados.....	29
Resposta de Modelos Lineares Generalizados.....	31
Preditores de Modelos Lineares Generalizados.....	32
Modelo de Modelos Lineares Generalizados.....	33
Estimação de Modelos Lineares Generalizados.....	34
Estatísticas de Modelos Lineares Generalizados.....	35
Médias de EM de Modelos Lineares Generalizados.....	36
Salvamento de Modelos Lineares Generalizados.....	37
Exportação de Modelos Lineares Generalizados.....	38
Variáveis Adicionais de Comando do GENLIN.....	39
Equações de estimativa generalizada.....	39
Tipo de Modelo de Equações de Estimativa Generalizada.....	41
Resposta de Equações de Estimativa Generalizada.....	43
Preditores de Equações de Estimativa Generalizada.....	44
Modelo de Equações de Estimativa Generalizada.....	44

Estimação de Equações de Estimativa Generalizada.....	45
Estatísticas de Equações de Estimativa Generalizada.....	47
Médias de EM de Equações de Estimativa Generalizada.....	48
Salvamento de Equações de Estimativa Generalizada.....	49
Exportação de Equações de Estimativa Generalizada.....	49
Variáveis Adicionais de Comando do GENLIN.....	50
Modelos lineares generalizados mistos.....	50
Obtendo um modelo linear generalizado misto.....	53
Resposta .....	53
Efeitos Fixos .....	55
Efeitos Aleatórios .....	56
Ponderação e Offset .....	57
Opções de Criação Geral .....	58
Estimação .....	58
Médias Estimadas .....	59
Salvar .....	60
Exportar .....	60
Visualização do modelo .....	61
Análise de log linear de seleção do modelo.....	65
Análise Log-Linear - Definir Intervalo.....	65
Modelo de Análise Log-Linear.....	66
Opções de Análise Log-Linear de Seleção de Modelo.....	66
Recursos Adicionais do Comando HILOGLINEAR.....	66
Análise de log linear geral.....	67
Modelo de Análise Log-Linear Geral.....	68
Opções de Análise Log-Linear Geral.....	68
Salvamento de Análise Log-Linear Geral.....	69
Recursos Adicionais do Comando GENLOG.....	69
Análise Log-Linear Logit.....	69
Modelo de Análise Log-Linear Logit.....	71
Opções de Análise Log-Linear Logit.....	71
Salvamento de Análise Log-Linear Logit.....	72
Recursos Adicionais do Comando GENLOG.....	72
Tabelas de mortalidade.....	72
Definir Intervalo para Tabelas de Mortalidade.....	74
Opções de Tabelas de Mortalidade.....	74
Variáveis Adicionais de Comando do SURVIVAL.....	74
Modelos De Tempo De Falha Acelerada Paramétrica.....	74
Modelos De Tempo De Falha Acelerada Paramétrica: Critérios.....	75
Modelos De Tempo De Falha Acelerada Paramétrica: Modelo.....	76
Modelos De Tempo De Falha Acelerada Paramétrica: Estimativa.....	76
Modelos De Tempo De Falha Acelerada Paramétrica: Imprimir.....	78
Modelos De Tempo De Falha Acelerada Paramétrica: Prever.....	78
Modelos De Tempo De Falha Acelerada Paramétrica: Plot.....	79
Modelos De Tempo De Falha Acelerada Paramétrica: Exportação.....	80
Sobrevivência AFT Definir Eventos para Variáveis de Status.....	80
Modelos de Tempo de Falha Acelerada Paramétrica: Selecionar Categoria.....	80
Modelos de fragilidade compartilhados paramétricos.....	81
Modelos de Fragilidade Compartilhados Paramétricos: Critérios.....	82
Modelos de fragilidade compartilhados paramétricos: Modelo.....	83
Modelos Paramétricos de Fragilidade Compartilhados: Estimativa.....	83
Modelos de fragilidade compartilhados paramétricos: Impressão.....	85
Modelos de fragilidade compartilhados paramétricos: prever.....	85
Modelos de Fragilidade Compartilhados Paramétricos: Plot.....	86
Modelos paramétricos compartilhados de fragilidade: Exportação.....	87
Modelos de fragilidade compartilhados paramétricos: definir eventos.....	87
Modelos de fragilidade compartilhados paramétricos-Exemplos.....	88

Modelos paramétricos compartilhados de fragilidade-um estudo de caso para dados recuperados.....	90
Análise de Sobrevivência de Kaplan-Meier.....	92
Kaplan-Meier - Definir Evento para Variável de Status.....	93
Kaplan-Meier - Comparar Níveis do Fator.....	93
Kaplan-Meier - Salvar Novas Variáveis.....	93
Opções de Kaplan-Meier.....	94
Variáveis Adicionais de Comando do KM.....	94
Análise de Regressão de Cox.....	94
Regressão de Cox - Definir Variáveis Categóricas.....	95
Gráficos de Regressão de Cox.....	96
Regressão de Cox - Salvar Novas Variáveis.....	96
Opções de Regressão de Cox.....	97
Regressão de Cox - Definir Evento para Variável de Status.....	97
Variáveis Adicionais de Comando do COXREG.....	97
Calculando Covariáveis Dependentes de Tempo.....	98
Calculando uma Covariável Dependente de Tempo.....	98
Esquemas de codificação de variável categórica.....	99
Desvio.....	99
Simples.....	99
Helmert.....	100
Diferença.....	100
Polinomial.....	100
Repetido.....	101
Especial.....	101
Indicador.....	102
Estruturas de Covariâncias.....	102
Estatísticas bayesianas.....	106
Inferência de uma amostra bayesiana: normal.....	107
Inferência de uma amostra bayesiana: binômio.....	111
Inferência de uma amostra bayesiana: Poisson.....	113
Inferência de amostra relacionada bayesiana: normal.....	115
Inferência de amostra independente bayesiana.....	116
Inferência bayesiana sobre correlação de Pearson.....	121
Inferência bayesiana sobre Modelos de regressão linear.....	123
ANOVA Bayesiano de um fator.....	128
Modelos Log-linear bayesianos.....	131
Modelos de ANOVA de medidas repetidas de um fator bayesiana.....	134
Regressão de Kernel Ridge.....	136
Parâmetros de kernel.....	138
Regressão de Kernel Ridge: opções.....	138
<b>Avisos.....</b>	<b>141</b>
Marcas comerciais.....	142
<b>Índice remissivo.....</b>	<b>145</b>



---

# Capítulo 1. Estatísticas avançadas

Os recursos de estatísticas avançadas a seguir estão incluídos em SPSS Statistics Edição padrão ou na opção Estatísticas avançadas.

---

## Introdução às Estatísticas Avançadas

A opção Estatística Avançada inclui procedimentos que oferecem opções de modelagem mais avançadas do que estão disponíveis através da opção Base de Estatística.

- O GLM Multivariate estende o modelo linear geral fornecido pelo GLM Univariate para permitir diversas variáveis dependentes. Uma extensão adicional, o GLM Repeated Measures, permite medições repetidas de diversas variáveis dependentes.
- A Análise de Componentes de Variância é uma ferramenta específica para decompor a variabilidade de uma variável dependente em componentes fixos e aleatórios.
- Os Modelos Lineares Mistos expandem o modelo linear geral para que os dados possam exibir a variabilidade correlacionada e não constante. O modelo linear misto, portanto, fornece a flexibilidade de modelar não apenas as médias dos dados, mas também as variâncias e covariâncias.
- Os Modelos Lineares Generalizados (GZLM) afrouxam a suposição de normalidade para o termo de erro e requerem apenas que a variável dependente esteja linearmente relacionada aos preditores por meio de uma função de transformação ou de ligação. As Equações de Estimativa Generalizada (GEE) estendem os GZLM para permitir medições repetidas.
- A Análise Log-Linear Geral permite ajustar modelos para dados de contagem de classificação cruzada, e a Análise Log-Linear de Seleção de Modelo pode ajudar a escolher entre os modelos.
- A Análise Log-Linear Logit permite ajustar modelos log-lineares para análise do relacionamento entre um dependente categórico e um ou mais preditores categóricos.
- A análise de sobrevivência está disponível por meio de Tabelas de Vida, para examinar a distribuição de variáveis de sobrevivência, possivelmente por níveis de uma variável de fator; por meio de Análise de Sobrevivência Kaplan-Meier, para examinar a distribuição de variáveis de sobrevivência, possivelmente por níveis de uma variável de fator ou produzindo análises separadas por níveis de uma variável de estratificação; e por meio de Regressão de Cox, para modelar o tempo para um evento especificado com base nos valores de covariáveis fornecidos.

---

## Análise do GLM Multivariado

O procedimento do GLM Multivariado fornece análise de regressão e análise de variância para múltiplas variáveis dependentes por uma ou mais variáveis ou covariáveis de fator. As variáveis de fator dividem a população em grupos. Usando esse procedimento de modelo linear geral, é possível testar hipóteses nulas sobre os efeitos de variáveis de fator nas médias de vários agrupamentos de uma distribuição em conjunto de variáveis dependentes. É possível investigar as interações entre fatores, bem como os efeitos de fatores individuais. Além disso, os efeitos de covariáveis e de interações de covariáveis com fatores podem ser incluídos. Para análise de regressão, as variáveis independentes (preditores) 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. Em um modelo multivariado, as somas de quadrados em razão dos efeitos nas somas de quadrados de modelo e erro estão em formulário matriz ao invés de no formulário escalar localizado em análise univariada. Essas matrizes são chamadas matrizes de SSCP (somas de quadrados e produtos cruzados). Se mais de uma variável dependente for especificada, a análise multivariada de variância usando rastreio de Pillai, lambda de Wilks, rastreio de Hotelling e critério maior de raiz de Roy com estatística  $F$  aproximada é fornecida, assim como a análise univariada de variância para cada variável dependente. Além de testar hipóteses, o GLM Multivariado produz estimativas paramétricas.

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 de valores médios preditos para as células no modelo e gráficos de perfil (gráficos de interação) dessas médias permitem visualizar facilmente alguns dos relacionamentos. Os testes de comparação múltiplos a posteriori são executados para cada variável dependente separadamente.

Valores residuais, valores preditos, a distância de Cook e valores de ponto de alavanca podem ser salvos como novas variáveis no arquivo de dados para verificação de suposições. Também estão disponíveis uma matriz de SSCP residual, que é uma matriz quadrada de somas de quadrados e produtos cruzados de resíduos, uma matriz de covariância residual, que é a matriz de SSCP residual dividida pelos graus de liberdade dos resíduos e a matriz de correlações residuais, que é a forma padronizada da matriz de covariância residual.

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 de medição diferente.

**Exemplo.** Um fabricante de plásticos mede três propriedades de filme plástico: resistência a rasgo, brilho e opacidade. Duas taxas de extrusão e duas quantias diferentes de aditivo são tentadas e as três propriedades são medidas sob cada combinação de taxa de extrusão e quantia de aditivo. O fabricante considera que a taxa de extrusão e a quantia de aditivo produzem individualmente resultados significativos, mas que a interação dos dois fatores não é significativa.

**Métodos.** Somas dos quadrados do Tipo I, Tipo II, Tipo III e Tipo IV podem ser usadas para avaliar hipóteses diferentes. O Tipo III é o padrão.

**Estatísticas.** Testes de intervalo e comparações múltiplas a posteriori: diferença menos significativa, Bonferroni, Sidak, Scheffé, *F* múltiplo de Ryan-Einot-Gabriel-Welsch, 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. Estatísticas descritivas: médias observadas, desvios padrão e contagens para todas as variáveis dependentes em todas as células o teste de homogeneidade de variâncias de Levene; teste de homogeneidade *M* de Caixa das matrizes de covariâncias das variáveis dependentes; e teste de esfericidade de Bartlett.

**Gráficos.** Dispersão versus nível, de resíduo e perfil (interação).

Considerações de Dados de GLM Multivariado

**Dados.** As variáveis dependentes devem ser quantitativas. Fatores são categóricos e podem ter valores numéricos ou valores da sequência de caracteres. Covariáveis são variáveis quantitativas que estão relacionadas à variável dependente.

**Suposições.** Para variáveis dependentes, os dados são uma amostra aleatória de vetores a partir de uma população normal multivariada; na população, as matrizes de variância-covariância para todas as células 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 gráficos de testes de homogeneidade de variâncias (incluído *M* de Caixa) e de dispersão versus nível. Também é possível examinar residuais e gráficos de resíduos.

**Procedimentos relacionados.** Use o procedimento Explorar para examinar os dados antes de fazer uma análise de variância. Para uma única variável dependente, use GLM Univariado. Se você mediu as mesmas variáveis dependentes em diversas ocasiões para cada assunto, use Medidas Repetidas do GLM.

Obtendo Tabelas do GLM Multivariado

1. Nos menus, escolha:

**Analisar > Modelo Linear Geral > Multivariado...**

2. Selecione pelo menos duas variáveis dependentes.

Opcionalmente, é possível especificar Fator(es) Fixo(s), Covariável(is) e Ponderação de WLS.



## Modelo GLM Multivariate

**Especificar modelo.** Um modelo fatorial completo contém todos os principais efeitos do fator, todos os principais efeitos covariáveis e todas as interações fator por fator. Ele não contém interações 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. Após selecionar **Customizado**, é possível selecionar os efeitos e interações principais que forem de interesse em sua análise.

**Soma dos quadrados.** O método de cálculo das 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 for possível presumir que os dados percorrerão a origem, será possível excluir o intercepto.

## Termos de construção e termos customizados

### Construir termos

Use esta opção quando quiser incluir termos não aninhados de um tipo específico (como efeitos principais) para todas as combinações de um conjunto selecionado de fatores e covariáveis.

### Construir termos customizados

Use esta opção quando quiser incluir termos aninhados ou quando quiser construir explicitamente qualquer variável de termo por variável. A construção de um termo aninhado envolve as seguintes etapas:

## Soma dos Quadrados

Para o modelo, é possível escolher um tipo de somas dos quadrados. O Tipo III é o utilizado com mais frequência e é o padrão.

**Tipo I.** Esse método também é conhecido como decomposição hierárquica do método da soma dos quadrados. Cada termo é ajustado apenas para o termo que o precede no modelo. As somas dos quadrados do Tipo I são normalmente utilizadas para:

- Um modelo ANOVA balanceado, no qual quaisquer efeitos principais serão especificados antes de quaisquer efeitos de interação de primeira ordem, quaisquer efeitos de interação de primeira ordem são especificados antes de quaisquer efeitos de interação de segunda ordem, e assim por diante.
- Um modelo de regressão polinomial no qual quaisquer termos de ordem inferior são especificados antes de quaisquer termos de ordem superior.
- Um modelo puramente aninhado em que o primeiro efeito especificado é aninhado dentro do segundo efeito especificado, o segundo efeito especificado é aninhado dentro do terceiro, e assim por diante. (Esta forma de aninhamento pode ser especificada apenas utilizando a sintaxe).

**Tipo II.** Este método calcula a soma 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 da soma dos quadrados do Tipo II é comumente utilizado 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 a soma dos quadrados de um efeito no design como somas de quadrados, ajustado para quaisquer outros efeitos que não contêm o efeito, e ortogonal para quaisquer efeitos (se houver) que contêm o efeito. As somas dos quadrados do Tipo III possuem uma vantagem maior por serem invariáveis com relação às frequências de célula, desde 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 de médias dos quadrados ponderados de Yates. O método da soma dos quadrados do Tipo III é normalmente utilizado para:

- Todos os modelos listados em Tipo I e Tipo II.
- Qualquer modelo balanceado ou não balanceado sem células vazias.

**Tipo IV.** Este método é projetado para uma situação em que houver células omissas. 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 da soma dos quadrados do Tipo IV é comumente utilizado para:

- Todos os modelos listados em Tipo I e Tipo II.
- Qualquer modelo balanceado ou não balanceado com células vazias.

## Contrastes do GLM Multivariate

Contrastes são utilizados para testar se os níveis de um efeito são significativamente diferentes entre si. É possível especificar um contraste para cada fator no modelo. Os contrastes representam as combinações lineares dos parâmetros.

Os testes de hipótese baseiam-se na hipótese nula  $\mathbf{LBM} = \mathbf{0}$ , em que  $\mathbf{L}$  é a matriz de coeficientes de contraste,  $\mathbf{M}$  é a matriz de identidade (que possui dimensão igual ao número de variáveis dependentes) e  $\mathbf{B}$  é o vetor do parâmetro. Quando um contraste é especificado, uma matriz  $\mathbf{L}$  é criada de forma que as colunas correspondentes ao fator correspondam ao contraste. As colunas restantes são ajustadas para que a matriz  $\mathbf{L}$  seja estimável.

Além do teste univariado que utiliza estatísticas  $F$  e os intervalos de confiança simultâneos do tipo Bonferroni com base na distribuição  $t$  de Student para as diferenças de contraste em todas as variáveis dependentes, são fornecidos testes multivariados que utilizam rastreio de Pillai, lambda de Wilks, rastreio de Hotelling e critérios raízes maiores de Roy.

Os contrastes disponíveis são desvio, simples, diferença, Helmert, repetido e polinomial. Para contrastes de desvio e contrastes simples, é possível escolher se a categoria de referência será a primeira ou a última 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 a referência.

**Diferença.** Compara a média de cada nível (exceto do primeiro) com a média dos níveis anteriores. (Às vezes chamada de contrastes de Helmert reversos.)

**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 do GLM Multivariate

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 no qual 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 estão disponíveis para gráficos. Os gráficos de perfil são criados para cada variável dependente.

Um gráfico de perfil de um fator mostra se as médias marginais estimadas estão aumentando ou diminuindo entre níveis. Para dois ou mais fatores, linhas paralelas indicam que não existe interação entre fatores, o que significa que é possível investigar os níveis de apenas um fator. Linhas não paralelas indicam uma interação.

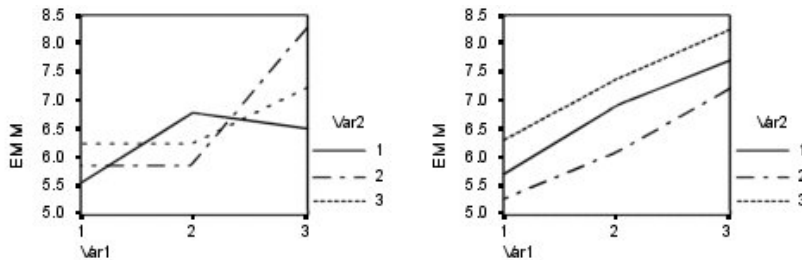


Figura 1. Gráfico não paralelo (esquerda) e gráfico paralelo (direita)

Depois de um gráfico ser especificado selecionando os fatores para o eixo horizontal e, opcionalmente, os fatores para linhas separadas e gráficos separados, o gráfico deve ser incluído na lista Gráficos.

## Comparações Post Hoc do GLM Multivariate

**Testes de comparação múltipla post hoc.** Após determinar que diferenças existem entre as médias, testes de intervalo post hoc e comparações múltiplas entre pares podem determinar quais médias diferem. As comparações são feitas em valores sem ajuste. Os testes post hoc são executados para cada variável dependente separadamente.

Os testes da diferença significativa honesta de Bonferroni e Tukey geralmente são usados por 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 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 da diferença significativa honesta 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 usado. Geralmente, o teste de Tukey é mais poderoso. O **teste de comparação entre pares de Gabriel** também usa o módulo máximo estudentizado e geralmente é 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.

**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 a categoria de controle) do fator não é igual à da categoria de controle, use um teste de dois lados. 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 intervalo de redução múltipla. Os procedimentos de redução múltipla primeiro testam se todas as médias são iguais. Se todas as médias

não forem iguais, os subconjuntos de médias serão testadas para 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 eles não são recomendados para os tamanhos de células desiguais.

Quando as variâncias forem desiguais, use **T2 de Tamhane** (teste de comparação entre pares conservador com base em um teste  $t$ ), **T3 de Dunnett** (teste de comparação entre pares com base no Módulo máximo estudentizado), **teste de comparação entre pares Games-Howell** (às vezes, liberals) ou o **C de Dunnett** (teste de comparação entre pares com base no intervalo estudentizado).

O **teste de amplitude múltipla de Duncan**, Student-Newman-Keuls (**S-N-K**) e o **b de Tukey** são testes de intervalo que classificam médias de grupo e calculam um valor de intervalo. Esses testes não são usados tão frequentemente quanto os testes discutidos anteriormente.

O **teste t de Waller-Duncan** usa uma abordagem bayesiana. Esse teste de intervalo usa a média harmônica do tamanho da amostra quando os tamanhos da amostra 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 geralmente mais conservador que outros testes, o que significa que uma diferença maior entre as médias é necessária para obter 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 será 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. Os subconjuntos homogêneos para testes de intervalo são fornecidas para S-N-K,  $b$  de Tukey, Duncan, R-E-G-W  $F$ , R-E-G-W  $Q$  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.

## Médias marginais estimadas de GLM

Selecione os fatores e interações para os quais deseja estimar as médias marginais da população nas células. Essas médias são ajustadas para as covariáveis, se houver.

### Comparar efeitos principais

Fornece comparações entre pares não corrigidas entre as médias marginais estimadas para qualquer efeito principal no modelo, para fatores entre e dentro-sujeitos. Esse item estará disponível somente se os efeitos principais forem selecionados na lista Médias de Exibição.

### Comparar efeitos principais simples

A configuração é ativada sempre que a lista de destino contém um ou mais efeitos de produto ou interação (por exemplo,  $A \times B$ ,  $A \times B \times C$ ). 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 o ajustamento de diferença menos significativa (LSD), Bonferroni ou Sidak ou 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**.

## Salvamento de GLM

É possível salvar valores preditos pelo modelo, resíduos e medidas relacionadas como novas variáveis no Editor de Dados. Muitas dessas variáveis podem ser usadas para examinar suposições sobre os dados.

Para salvar os valores para usar em outra sessão do IBM SPSS Estatísticas, 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, estudantizados e excluídos também estão disponíveis. Se uma variável WLS foi escolhida, resíduos não padronizados ponderados estarão 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 Estatísticas. 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 as estatísticas consistentes de Heteroskedasticity são selecionadas (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 leem arquivos de matriz.

## Opções do GLM Multivariado

Estatísticas de opcionais estão disponíveis a partir dessa caixa de diálogo. As estatísticas são calculadas utilizando um modelo de efeitos fixos.

**Exibição.** Selecione **Estatística descritiva** 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  $\eta^2$  quadrado parcial para cada efeito e cada estimativa paramétrica. A estatística  $\eta^2$  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 de parâmetros** para produzir as estimativas de parâmetros, os erros padrão, testes *t*, intervalos de confiança e a potência observada para cada teste. É possível exibir as **Matrizes**

de **SSCP** de hipótese e erros e a **Matriz de SSCP de resíduo** mais o teste de esfericidade de Bartlett da matriz de covariância residual.

Os **Testes de Homogeneidade** produzem o teste de Levene da homogeneidade da variância para cada variável dependente em todas as combinações de nível somente para os fatores entre assuntos. Além disso, testes de homogeneidade incluem teste de homogeneidade *M* de Caixa das matrizes de covariâncias das variáveis dependentes em todas as combinações de nível dos fatores entre assuntos. As opções de gráficos de dispersão versus nível e de resíduos são úteis para verificar suposições sobre os dados. Esse item fica desativado se não houver fatores. Selecione **Gráficos de resíduos** para produzir um gráfico de resíduos observado por predito por padronizado para cada variável dependente. Esses gráficos são úteis para investigar a suposição de variância igual. Selecione **Falta de teste 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.

## Exibir

### Estatística descritiva

Produz 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 estimativa de efeito e de parâmetro. A estatística eta quadrado descreve a proporção da variabilidade total atribuível a um fator.

### Potência observada

Obtém a potência do teste quando a hipótese alternativa é configurada com base no valor observado.

### Estimativas de parâmetro

Produz as estimativas de parâmetros, erros padrão, testes *t*, intervalos de confiança e a potência observada para cada teste.

### Matrizes SSCP

Exibe as matrizes SSCP de hipótese e de erro.

### Matriz SSCP de resíduo

Exibe a matriz SSCP residual de hipótese e de erro.

### Matriz de transformação

Exibe o teste de esfericidade de Bartlett da matriz de covariância residual.

### Testes de homogeneidade

Produz o teste de Levene da homogeneidade de variância para cada variável dependente em todas as combinações de níveis dos fatores entre assuntos, somente para fatores entre assuntos. Além disso, testes de homogeneidade incluem teste de homogeneidade *M* de Caixa das matrizes de covariâncias das variáveis dependentes em todas as combinações de nível dos fatores entre assuntos.

### Gráfico de nível vs. difusão

Útil para verificar suposições sobre os dados para investigar a suposição de variância igual. Esse item fica desativado se não houver fatores.

### Gráfico de resíduos

Produz um gráfico de resíduos padronizado observado por e padronizado por para cada variável dependente. O gráfico é útil para investigar a suposição de variância igual.

### Falta de ajuste

Verifique 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.

### Nível de significância

Talvez você queira ajustar o nível de significância utilizado em testes post hoc e o nível de confiança utilizado para construir intervalos de confiança. O valor especificado é também usado para calcular a potência observada para o teste. Ao especificar um nível de significância, o nível associado dos intervalos de confiança é exibido no diálogo.

## Recursos Adicionais de Comando do GLM

Esses recursos podem se aplicar à análise de medidas univariadas, multivariadas ou repetidas. 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 ou 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 da sintaxe de comando* para obter informações de sintaxe completa.

## GLM Repeated Measures

O procedimento de Medidas Repetidas do GLM fornece análise de variância quando a mesma medição é feita diversas vezes em cada assunto ou caso. Se fatores entre assuntos forem especificados, eles dividem a população em grupos. Usando esse procedimento de modelo linear geral, é possível testar hipóteses nulas sobre os efeitos de fatores entre sujeitos e fatores dentro-sujeitos. É possível investigar as interações entre fatores, bem como os efeitos de fatores individuais. Além disso, os efeitos de covariáveis constantes e de interações de covariáveis com os fatores entre assuntos podem ser incluídos.

Em um design de medidas repetidas duplamente multivariadas, as variáveis dependentes representam medições de mais de uma variável para os níveis diferentes de fatores entre assuntos. Por exemplo, você poderia ter medido pulso e respiração em três vezes diferentes em cada sujeito.

O procedimento de Medidas Repetidas do GLM fornece análises univariadas e multivariadas para os dados de medidas repetidas. Modelos balanceados e não balanceados podem ser testados. Um design será balanceado se cada célula no modelo contiver o mesmo número de casos. Em um modelo multivariado, as somas de quadrados em razão dos efeitos nas somas de quadrados de modelo e erro estão em formulário matriz ao invés de no formulário escalar localizado em análise univariada. Essas matrizes são chamadas matrizes de SSCP (somas de quadrados e produtos cruzados). Além de testar hipóteses, as Medidas Repetidas do GLM produzem estimativas paramétricas.

Contrastes *a priori* normalmente usados estão disponíveis para executar teste de hipótese em fatores entre assuntos. 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 de valores médios preditos para as células no modelo e gráficos de perfil (gráficos de interação) dessas médias permitem visualizar facilmente alguns dos relacionamentos.

Valores residuais, valores preditos, a distância de Cook e valores de ponto de alavanca podem ser salvos como novas variáveis no arquivo de dados para verificação de suposições. Também estão disponíveis uma matriz de SSCP residual, que é uma matriz quadrada de somas de quadrados e produtos cruzados de resíduos, uma matriz de covariância residual, que é a matriz de SSCP residual dividida pelos graus de liberdade dos resíduos e a matriz de correlações residuais, que é a forma padronizada da matriz de covariância residual.

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 de medição diferente.

**Exemplo.** Doze alunos são designados para um grupo de alta ou baixa ansiedade com base em seus escores em um teste de classificação de ansiedade. A classificação de ansiedade é chamada de fator entre assuntos porque ela divide os assuntos em grupos. São fornecidas quatro avaliações aos alunos em uma tarefa de aprendizado e o número de erros para cada avaliação é registrado. Os erros para cada avaliação são registrados em variáveis separadas e um fator (uma avaliação) dentre-sujeitos é definido com quatro níveis para as quatro avaliações. O efeito da avaliação é considerado significativo, enquanto a interação de avaliação por ansiedade não é significativa.

**Métodos.** Somas dos quadrados do Tipo I, Tipo II, Tipo III e Tipo IV podem ser usadas para avaliar hipóteses diferentes. O Tipo III é o padrão.

**Estatísticas.** Testes de intervalo posteriori e múltiplas comparações (para fatores entre sujeitos): diferença menos significativa, Bonferroni, Sidak, Scheffé,  $F$  múltiplo de Ryan-Einot-Gabriel-Welsch, amplitude múltipla de Ryan-Einot-Gabriel-Welsch, diferença honestamente significativa de Student-Newman-Keuls, Tukey,  $b$  de Tukey, GT2 de Duncan, Hochberg, teste  $t$  de Waller Duncan, Dunnett (unilateral e bilateral), T2 de Tamhane, T3 de Dunnett, C de Games-Howell e Dunnett. Estatística descritiva: médias observadas, desvios padrão e contagens para todas as variáveis dependentes em todas as células; o teste de Levene para homogeneidade de variância;  $M$  de Box e teste de esfericidade de Mauchly.

**Gráficos.** Dispersão versus nível, de resíduo e perfil (interação).

Considerações de Dados de Medidas Repetidas do GLM

**Dados.** As variáveis dependentes devem ser quantitativas. Fatores entre sujeitos dividem a amostra em subgrupos discretos, como homem e mulher. Esses fatores são categóricos e podem ter valores numéricos ou valores da sequência de caracteres. Fatores dentre-sujeitos são definidos na caixa de diálogo Fator(es) de Definição de Medidas. Covariáveis são variáveis quantitativas que estão relacionadas à variável dependente. Para uma análise de medidas repetidas, estas devem permanecer constantes em cada nível de uma variável dentre-sujeitos.

O arquivo de dados deve conter um conjunto de variáveis para cada grupo de medições nos assuntos. O conjunto tem uma variável para cada repetição da medição dentro do grupo. Um fator dentre-sujeitos é definido para o grupo com o número de níveis igual ao número de repetições. Por exemplo, medições de ponderação poderiam ser feitas em dias diferentes. Se medições da mesma propriedade foram feitas em cinco dias, o fator dentre-sujeitos poderia ser especificado como *dia* com cinco níveis.

Para fatores dentre-sujeitos, o número de medições para cada assunto é igual ao produto do número de níveis de cada fator. Por exemplo, se medições foram feitas em três vezes diferentes a cada dia por quatro dias, o número total de medições é 12 para cada assunto. Os fatores dentre-sujeitos poderiam ser especificados como *dia*(4) e *vezes*(3).

**Suposições.** Um análise de medida repetida pode ser abordada de duas maneiras, univariada e multivariada.

A abordagem univariada (também conhecida como o gráfico dividido ou a abordagem de modelo misto) considera as variáveis dependentes como respostas para os níveis de fatores dentre-sujeitos. As medições em um assunto devem ser uma amostra de uma distribuição normal multivariada e as matrizes de variância-covariância são as mesmas ao longo das células formadas pelos efeitos entre-sujeitos.



Determinadas premissas são feitas na matriz de variância-covariância das variáveis dependentes. A validade da estatística  $F$  usada na abordagem univariada pode ser assegurada se a matriz de variância-covariância for circular na forma (Huynh e Mandeville, 1979).

Para testar essa suposição, o teste de esfericidade de Mauchly pode ser usado, que executa um teste de esfericidade na matriz de variância-covariância de uma variável dependente transformada ortonormalizada. O teste de Mauchly é exibido automaticamente para uma análise de medidas repetidas. Para tamanhos da amostra pequenos, esse teste não é muito eficaz. Para tamanhos da amostra grandes, o teste pode ser significativo mesmo quando o impacto da partida sobre os resultados for pequeno. Se a significância do teste for grande, a hipótese de esfericidade pode ser assumida. No entanto, se a significância for pequena e a suposição de esfericidade parecer estar violada, um ajustamento nos graus de liberdade do numerador e do denominador poderá ser feito para validar a estatística  $F$  univariada. Três estimativas desse ajustamento, que é chamado **epsilon**, estão disponíveis no procedimento de Medidas Repetidas do GLM. Os graus de liberdade do numerador e do denominador devem ser multiplicados por epsilon e a significância da razão  $F$  deve ser avaliada com os novos graus de liberdade.

A abordagem multivariada considera as medições em um assunto como uma amostra de uma distribuição normal multivariada e as matrizes de variância-covariância são as mesmas ao longo das células formadas pelos efeitos entre-sujeitos. Para testar se as matrizes de variância-covariância ao longo das células são as mesmas, o teste  $M$  de Caixa pode ser usado.

**Procedimentos relacionados.** Use o procedimento Explorar para examinar os dados antes de fazer uma análise de variância. Se *não* houver medidas repetidas em cada assunto, use GLM Univariado ou GLM Multivariado. Se houver somente duas medições para cada assunto (por exemplo, pré-teste e pós-teste) e não houver nenhum fator entre assuntos, é possível usar o procedimento de Teste T de Amostras Em Pares.

Obtendo Medidas Repetidas do GLM

1. Nos menus, escolha:

**Analisar > Modelo Linear Geral > Medidas Repetidas...**

2. Digite um nome do fator dentre-sujeitos e seu número de níveis.

3. Clique em **Incluir**.

4. Repita essas etapas para cada fator dentre-sujeitos.

Para definir fatores de medida para um design de medidas repetidas duplamente multivariadas:

5. Digite o nome da medida.

6. Clique em **Incluir**.

Após definir todos os seus fatores e medidas:

7. Clique em **Definir**.

8. Selecione uma variável dependente que corresponda a cada combinação de fatores dentre-sujeitos (e, opcionalmente, medidas) na lista.

Para mudar ranqueamentos das variáveis, use as setas para cima e para baixo.

Para fazer mudanças nos fatores dentre-sujeitos, é possível reabrir a caixa de diálogo Medidas Repetidas - Definir Fator(es) sem fechar a caixa de diálogo principal. Opcionalmente, é possível especificar fator(es) e covariáveis entre assuntos.

## Medidas Repetidas do GLM - Definir Fatores

Medidas Repetidas do GLM analisam grupos de variáveis dependentes relacionadas que representam medições diferentes do mesmo atributo. Essa caixa de diálogo permite que você defina um ou mais fatores dentre-sujeitos para usar nas Medidas Repetidas do GLM. Observe que a ordem na qual você especifica fatores dentre-sujeitos é importante. Cada fator constitui um nível dentro do fator anterior.

Para usar Medidas Repetidas, deve-se configurar os seus dados corretamente. Deve-se definir fatores dentre-sujeitos nessa caixa de diálogo. Observe que esses fatores não são variáveis existentes em seus dados, mas de preferência fatores que você define aqui.

**Exemplo.** Em um estudo de perda de ponderação, suponha que as ponderações de diversas pessoas sejam medidas a cada semana por cinco semanas. No arquivo de dados, cada pessoa é um assunto ou caso. As ponderações para as semanas são registradas nas variáveis *ponderação1*, *ponderação2*, e assim por diante. O sexo de cada pessoa é registrado em uma outra variável. As ponderações, medidas para cada assunto repetidamente, podem ser agrupadas definindo um fator dentre-sujeitos. O fator poderia ser chamado de *semana*, definido para ter cinco níveis. Na caixa de diálogo principal, as variáveis *weight1*, ..., *weight5* são usados para atribuir os cinco níveis de *semana*. A variável no arquivo de dados que agrupa homens e mulheres (*sexo*) pode ser especificada como fator entre assuntos para estudar as diferenças entre homens e mulheres.

**Medidas.** Se assuntos foram testados em mais de uma medida a cada vez, defina as medidas. Por exemplo, a taxa de pulso e respiração poderia ser medida em cada sujeito todos os dias por uma semana. Essas medidas não existem como variáveis no arquivo de dados, mas são definidas aqui. Um modelo com mais de uma medida é às vezes chamado de modelo de medidas repetidas duplamente multivariadas.

## Modelo de Medidas Repetidas do GLM

**Especificar modelo.** Um modelo fatorial completo contém todos os principais efeitos do fator, todos os principais efeitos covariáveis e todas as interações fator por fator. Ele não contém interações 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.

**Entre Assuntos.** Os fatores e covariáveis entre assuntos são listados.

**Modelo.** O modelo depende da natureza de seus dados. Após selecionar **Customizado**, é possível selecionar os efeitos e as interações dentre-sujeitos e os efeitos e as interações entre assuntos que são de interesse em sua análise.

**Soma dos quadrados.** O método de cálculo das somas dos quadrados para o modelo entre assuntos. Para modelos balanceados ou não balanceados entre assuntos sem nenhuma célula omissa, o método da soma dos quadrados do Tipo III é o mais comumente usado.

## Termos de construção e termos customizados

### Construir termos

Use esta opção quando quiser incluir termos não aninhados de um tipo específico (como efeitos principais) para todas as combinações de um conjunto selecionado de fatores e covariáveis.

### Construir termos customizados

Use esta opção quando quiser incluir termos aninhados ou quando quiser construir explicitamente qualquer variável de termo por variável. A construção de um termo aninhado envolve as seguintes etapas:

## Soma dos Quadrados

Para o modelo, é possível escolher um tipo de somas dos quadrados. O Tipo III é o utilizado com mais frequência e é o padrão.

**Tipo I.** Esse método também é conhecido como decomposição hierárquica do método da soma dos quadrados. Cada termo é ajustado apenas para o termo que o precede no modelo. As somas dos quadrados do Tipo I são normalmente utilizadas para:

- Um modelo ANOVA balanceado, no qual quaisquer efeitos principais serão especificados antes de quaisquer efeitos de interação de primeira ordem, quaisquer efeitos de interação de primeira ordem são especificados antes de quaisquer efeitos de interação de segunda ordem, e assim por diante.
- Um modelo de regressão polinomial no qual quaisquer termos de ordem inferior são especificados antes de quaisquer termos de ordem superior.
- Um modelo puramente aninhado em que o primeiro efeito especificado é aninhado dentro do segundo efeito especificado, o segundo efeito especificado é aninhado dentro do terceiro, e assim por diante. (Esta forma de aninhamento pode ser especificada apenas utilizando a sintaxe).

**Tipo II.** Este método calcula a soma 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 da soma dos quadrados do Tipo II é comumente utilizado 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 a soma dos quadrados de um efeito no design como somas de quadrados, ajustado para quaisquer outros efeitos que não contêm o efeito, e ortogonal para quaisquer efeitos (se houver) que contêm o efeito. As somas dos quadrados do Tipo III possuem uma vantagem maior por serem invariáveis com relação às frequências de célula, desde 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 de médias dos quadrados ponderados de Yates. O método da soma dos quadrados do Tipo III é normalmente utilizado para:

- Todos os modelos listados em Tipo I e Tipo II.
- Qualquer modelo balanceado ou não balanceado sem células vazias.

**Tipo IV.** Este método é projetado para uma situação em que houver células omissas. 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 da soma dos quadrados do Tipo IV é comumente utilizado para:

- Todos os modelos listados em Tipo I e Tipo II.
- Qualquer modelo balanceado ou não balanceado com células vazias.

## Contrastes do GLM Repeated Measures

Contrastes são utilizados para testar para as diferenças entre os níveis de um fator entre assuntos. É possível especificar um contraste para cada fator entre assuntos. Os contrastes representam as combinações lineares dos parâmetros.

Os testes de hipótese baseiam-se na hipótese nula  $\mathbf{LBM} = 0$ , em que  $\mathbf{L}$  é a matriz de coeficientes de contraste,  $\mathbf{B}$  é o vetor paramétrica e  $\mathbf{M}$  é a matriz da média que corresponde à transformação média para a variável dependente. É possível exibir essa matriz de transformação ao selecionar **Matriz de Transformação** na caixa de diálogo Opções de Medidas Repetidas. Por exemplo, se quatro variáveis dependentes, um fator dentre-sujeitos de quatro níveis e contraste polinomiais (o padrão) forem utilizados para fatores dentre-sujeitos, a matriz  $\mathbf{M}$  será  $(0,5 \ 0,5 \ 0,5 \ 0,5)'$ . Quando um contraste é especificado, uma matriz  $\mathbf{L}$  é criada de forma que as colunas correspondentes ao fator entre assuntos correspondam ao contraste. As colunas restantes são ajustadas para que a matriz  $\mathbf{L}$  seja estimável.

Os contrastes disponíveis são desvio, simples, diferença, Helmert, repetido e polinomial. Para contrastes de desvio e contrastes simples, é possível escolher se a categoria de referência será a primeira ou a última categoria.

Um contraste diferente de **Nenhum** deve ser selecionado para fatores dentre-sujeitos.

## 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 a referência.

**Diferença.** Compara a média de cada nível (exceto do primeiro) com a média dos níveis anteriores. (Às vezes chamada de contrastes de Helmert reversos.)

**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.

## Contrastes de Perfil do GLM Repeated Measures

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 no qual 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 estão disponíveis para gráficos. Os gráficos de perfil são criados para cada variável dependente. Os fatores entre assuntos e fatores dentro-sujeitos podem ser utilizados em gráficos de perfil.

Um gráfico de perfil de um fator mostra se as médias marginais estimadas estão aumentando ou diminuindo entre níveis. Para dois ou mais fatores, linhas paralelas indicam que não existe interação entre fatores, o que significa que é possível investigar os níveis de apenas um fator. Linhas não paralelas indicam uma interação.

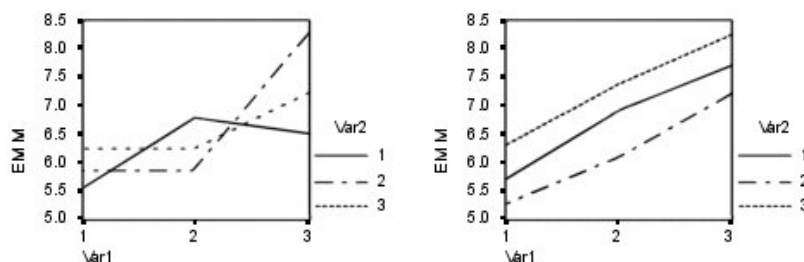


Figura 2. Gráfico não paralelo (esquerda) e gráfico paralelo (direita)

Depois de um gráfico ser especificado selecionando os fatores para o eixo horizontal e, opcionalmente, os fatores para linhas separadas e gráficos separados, o gráfico deve ser incluído na lista Gráficos.

## Comparações Post Hoc do GLM Repeated Measures

**Testes de comparação múltipla post hoc.** Após determinar que diferenças existem entre as médias, testes de intervalo post hoc e comparações múltiplas entre pares podem determinar quais médias diferem. As comparações são feitas em valores sem ajuste. Esses testes não estarão disponíveis se não houver fatores entre assuntos, e os testes de comparação múltipla post hoc são executados para a média entre os níveis dos fatores dentro-sujeitos.

Os testes da diferença significativa honesta de Bonferroni e Tukey geralmente são usados por 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 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 da diferença significativa honesta 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 usado. Geralmente, o teste de Tukey é mais poderoso. O **teste de comparação entre pares de Gabriel** também usa o módulo máximo estudentizado e geralmente é 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.

**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 a categoria de controle) do fator não é igual à da categoria de controle, use um teste de dois lados. 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 intervalo de redução múltipla. Os procedimentos de redução múltipla primeiro testam se todas as médias são iguais. Se todas as médias não forem iguais, os subconjuntos de médias serão testadas para 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 eles não são recomendados para os tamanhos de células desiguais.

Quando as variâncias forem desiguais, use **T2 de Tamhane** (teste de comparação entre pares conservador com base em um teste  $t$ ), **T3 de Dunnett** (teste de comparação entre pares com base no Módulo máximo estudentizado), **teste de comparação entre pares Games-Howell** (às vezes, liberals) ou o **C de Dunnett** (teste de comparação entre pares com base no intervalo estudentizado).

O **teste de amplitude múltipla de Duncan**, Student-Newman-Keuls (**S-N-K**) e o **b de Tukey** são testes de intervalo que classificam médias de grupo e calculam um valor de intervalo. Esses testes não são usados tão frequentemente quanto os testes discutidos anteriormente.

O **teste t de Waller-Duncan** usa uma abordagem bayesiana. Esse teste de intervalo usa a média harmônica do tamanho da amostra quando os tamanhos da amostra 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 geralmente mais conservador que outros testes, o que significa que uma diferença maior entre as médias é necessária para obter 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 será 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. Os subconjuntos homogêneos para testes de intervalo são fornecidas para S-N-K,  $b$  de Tukey, Duncan, R-E-G-W  $F$ , R-E-G-W  $Q$  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.

## Médias marginais estimadas de GLM

Selecione os fatores e interações para os quais deseja estimar as médias marginais da população nas células. Essas médias são ajustadas para as covariáveis, se houver.

### Comparar efeitos principais

Fornece comparações entre pares não corrigidas entre as médias marginais estimadas para qualquer efeito principal no modelo, para fatores entre e dentro-sujeitos. Esse item estará disponível somente se os efeitos principais forem selecionados na lista Médias de Exibição.

### Comparar efeitos principais simples

A configuração é ativada sempre que a lista de destino contém um ou mais efeitos de produto ou interação (por exemplo,  $A \times B$ ,  $A \times B \times C$ ). 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 o ajustamento de diferença menos significativa (LSD), Bonferroni ou Sidak ou 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**.

## Salvamento de Medidas Repetidas do GLM

É possível salvar valores preditos pelo modelo, resíduos e medidas relacionadas como novas variáveis no Editor de Dados. Muitas dessas variáveis podem ser usadas para examinar suposições sobre os dados. Para salvar os valores para usar em outra sessão do IBM SPSS Estatísticas, 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.
- *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. Estão disponíveis valores de alavancagem de distância e não centralizados de Cook.

- *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, estudantizados e excluídos também estão disponíveis.

- *Não padronizado.* A diferença entre um valor observado e o valor predito 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 predito ajustado.

**Estatísticas de coeficiente.** Salva uma matriz de variância-covariância das estimativas paramétrica em um conjunto de dados ou arquivo de dados. Além disso, para cada variável dependente, haverá uma linha de estimativas paramétrica, uma linha de valores de significância para as estatísticas *t* correspondentes às estimativas paramétrica e uma linha de graus de liberdade de resíduos. Para um modelo multivariado, há linhas semelhantes para cada variável dependente. É possível usar esses dados de matriz em outros procedimentos que leem arquivos de matriz. 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.

## Opções de Medidas Repetidas do GLM

Estatísticas de opcionais estão disponíveis a partir dessa caixa de diálogo. As estatísticas são calculadas utilizando um modelo de efeitos fixos.

## Exibir

### Estatística descritiva

Produz 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 estimativa de efeito e de parâmetro. A estatística eta quadrado descreve a proporção da variabilidade total atribuível a um fator.

### Potência observada

Obtém a potência do teste quando a hipótese alternativa é configurada com base no valor observado.

### Estimativas de parâmetro

Produz as estimativas de parâmetros, erros padrão, testes  $t$ , intervalos de confiança e a potência observada para cada teste.

### Matrizes SSCP

Exibe as matrizes SSCP de hipótese e de erro.

### Matriz SSCP de resíduos

Exibe a matriz SSCP residual.

### Matriz de transformação

Exibe o teste de esfericidade de Bartlett da matriz de covariância residual.

### Testes de homogeneidade

Produz o teste de Levene da homogeneidade de variância para cada variável dependente em todas as combinações de níveis dos fatores entre assuntos, somente para fatores entre assuntos. Além disso, testes de homogeneidade incluem teste de homogeneidade  $M$  de Caixa das matrizes de covariâncias das variáveis dependentes em todas as combinações de nível dos fatores entre assuntos.

### Gráfico de nível vs. difusão

Útil para verificar suposições sobre os dados. Esta opção será desativada se não houver fatores.

### Gráfico de resíduos

Produz um gráfico de resíduos padronizado observado por e padronizado por para cada variável dependente. Esses gráficos são úteis para investigar a suposição de variância igual. Esta opção será desativada se não houver fatores.

### Falta de ajuste

Verifica 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.

### Nível de significância

Talvez você queira ajustar o nível de significância utilizado em testes post hoc e o nível de confiança utilizado para construir intervalos de confiança. O valor especificado é também usado para calcular a potência observada para o teste. Ao especificar um nível de significância, o nível associado dos intervalos de confiança é exibido no diálogo.

## Recursos Adicionais de Comando do GLM

Esses recursos podem se aplicar à análise de medidas univariadas, multivariadas ou repetidas. 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 da sintaxe de comando* para obter informações de sintaxe completa.

## Análise de Componentes de Variância

---

O procedimento Componente de Variância, para modelos de efeitos mistos, estima a contribuição de cada efeito aleatório com a variância da variável dependente. Esse procedimento é interessante principalmente para análise de modelos mistos, como gráfico de divisão, medidas repetidas univariadas e designs de bloco aleatório. Ao calcular componentes de variância, é possível determinar onde focar a atenção para reduzir a variância.

Quatro métodos diferentes estão disponíveis para estimar os componentes de variância: estimador imparcial quadrático de norma mínima (MINQUE), análise de variância (ANOVA), máxima verossimilhança (ML) e probabilidade máxima restrita (REML). Várias especificações estão disponíveis para métodos diferentes.

As saídas padrão para todos os métodos inclui estimativas de componente de variância. Se o método ML ou o método REML for utilizado, uma tabela de matrizes de covariância assintótica também será exibida. Outra saída disponível inclui uma tabela ANOVA e quadrados médios esperados para o método ANOVA e um histórico de iteração para os métodos ML e REML. O procedimento Componentes de Variância é totalmente compatível com o procedimento GLM Univariate.

A Ponderação de WLS permite especificar uma variável utilizada para atribuir às observações ponderações diferentes para uma análise ponderada, talvez para compensar variações na precisão de medição.

**Exemplo.** Em uma escola agrícola, os aumentos de peso de porcos em seis ninhadas diferentes são medidos após um mês. A variável de ninhada é um fator aleatório com seis níveis. (As seis ninhadas estudadas são uma amostra aleatória de uma grande população de ninhadas suínas). O investigador descobre que a variância no aumento de peso é muito mais atribuída à diferença nas ninhadas do que à diferença de suínos dentro de uma ninhada.

Considerações de Dados de Componentes de Variância

**Dados.** A variável dependente é quantitativa. Os fatores são categóricos. Eles podem ter valores numéricos ou valores de sequência de caracteres de até oito bytes. Pelo menos um dos fatores deve ser aleatório. Ou seja, os níveis do fator devem ser uma amostra aleatória de níveis possíveis. Covariáveis são variáveis quantitativas que estão relacionadas à variável dependente.

**Suposições.** Todos os métodos supõem que os parâmetros de modelo de um efeito aleatório possuem médias zeros, variâncias de constante finitas e são mutuamente não correlacionados. Os parâmetros do modelo de efeitos aleatórios diferentes também são não correlacionados.



O termo de resíduo também possui uma média zero e variância de constante finita. Ele não é correlacionado com os parâmetros de modelo de nenhum efeito aleatório. Termos de resíduo a partir de observações diferentes são assumidos como sendo não correlacionados.

Com base nessas suposições, as observações do mesmo nível de um fator aleatório são correlacionadas. Este fato distingue um modelo de componente de variância de um modelo linear geral.

O ANOVA e MINQUE não requerem suposições de normalidade. Eles são robustos para partidas moderadas da suposição de normalidade.

O ML e REML requerem que o parâmetro de modelo e o termo de resíduo sejam normalmente distribuídos.

**Procedimentos relacionados.** Utilize o procedimento Explorar para examinar os dados antes de fazer uma análise de componentes de variância. Para teste de hipótese, utilize GLM Univariate, GLM Multivariate e GLM Repeated Measures.

Obtendo Tabelas de Componentes de Variância

1. Nos menus, escolha:

**Analisar > Modelo Linear Geral > Componentes de Variância...**

2. Selecione uma variável dependente.

3. Selecione variáveis para Fator(es) Fixo(s), Fator(es) Aleatório(s) e Covariável(is), conforme apropriado para seus dados. Para especificar uma variável de ponderação, utilize Ponderação de WLS.

## Modelos de Componentes de Variância

**Especificar modelo.** Um modelo fatorial completo contém todos os principais efeitos do fator, todos os principais efeitos covariáveis e todas as interações fator por fator. Ele não contém interações 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 & covariáveis.** Os fatores e covariáveis são listados.

**Modelo.** O modelo depende da natureza de seus dados. Após selecionar **Customizado**, é possível selecionar os efeitos e interações principais que forem de interesse em sua análise. O modelo deve conter um fator aleatório.

Para os fatores e covariáveis selecionados:

### Interação

Cria o termo de interação de nível mais alto de todas as variáveis selecionadas. Esse é o padrão.

### Efeitos principais

Cria um termo dos principais efeitos para cada variável selecionada.

### Todos de 2 fatores

Cria todas as interações de dois fatores das variáveis selecionadas.

### Todos de 3 fatores

Cria todas as interações de três fatores possíveis das variáveis selecionadas.

### Todos de 4 fatores

Cria todas as interações de quatro fatores possíveis das variáveis selecionadas.

### Todos de 5 fatores

Cria todas as interações de cinco fatores possíveis das variáveis selecionadas.

**Incluir intercepto no modelo.** Normalmente o intercepto é incluído no modelo. Se for possível presumir que os dados percorrerão a origem, será possível excluir o intercepto.

## Termos de construção e termos customizados

### Construir termos

Use esta opção quando quiser incluir termos não aninhados de um tipo específico (como efeitos principais) para todas as combinações de um conjunto selecionado de fatores e covariáveis.

### Construir termos customizados

Use esta opção quando quiser incluir termos aninhados ou quando quiser construir explicitamente qualquer variável de termo por variável. A construção de um termo aninhado envolve as seguintes etapas:

## Opções de Componentes de Variância

**Método.** É possível escolher um dos quatro métodos para estimar os componentes de variância.

- **MINQUE** (estimador imparcial quadrático de norma mínima) produz estimativas que são invariáveis com relação aos efeitos fixos. Se os dados estiverem normalmente distribuídos e as estimativas corretas, esse método produzirá a variância mínima entre todos os estimadores imparciais. É possível escolher um método para ponderações informações a priori de efeito aleatório.
- **ANOVA** (análise de variância) calcula estimativas imparciais utilizando as somas dos quadrados de Tipo I ou Tipo III para cada efeito. Às vezes o método ANOVA produz estimativas de variância negativas, que podem indicar um modelo incorreto, um método de estimação impróprio ou a necessidade de mais dados.
- **Máxima verossimilhança** (ML) produz estimativas que seriam mais consistentes com os dados realmente observados, utilizando iterações. Essas estimativas podem ser tendenciosas. Esse método é assintoticamente normal. Estimativas de ML e REML são invariáveis na conversão. Este método não leva em conta os graus de liberdade usados para estimar os efeitos fixos.
- As estimativas **Máxima verossimilhança restrita** (REML) reduzem as estimativas ANOVA para muitos (se não todos) os casos de dados balanceados. Como esse método é ajustado para efeitos fixos, ele deverá ter erros padrão menores do que o método ML. Este método leva em conta os graus de liberdade usados para estimar os efeitos fixos.

**Informações a priori de Efeito Aleatório. Uniforme** implica que todos os efeitos aleatórios e o termo de resíduo possuem um impacto igual sobre as observações. O esquema **Zero** é equivalente a supor variâncias de efeito aleatório zero. Disponível somente para o método MINQUE.

**Soma dos Quadrados.** As somas dos quadrados do **Tipo I** são utilizadas para o modelo hierárquico, que é geralmente usado em literatura de componente de variância. Se escolher **Tipo III**, o padrão no GLM, as estimativas de variância poderão ser utilizadas no GLM Univariate para teste de hipótese com somas dos quadrados do Tipo III. Disponível somente para o método ANOVA.

**Critérios.** É possível especificar o critério de convergência e o número máximo de iterações. Disponível somente para os métodos ML ou REML.

**Exibição.** Para o método ANOVA, é possível optar por exibir somas de quadrados e quadrados médios esperados. Se **Máxima verossimilhança** ou **Máxima verossimilhança restrita** tiver sido selecionada, será possível exibir um histórico das iterações.

## Soma dos Quadrados (Componentes de Variância)

Para o modelo, é possível escolher um tipo de soma dos quadrados. O Tipo III é o utilizado com mais frequência e é o padrão.

**Tipo I.** Esse método também é conhecido como decomposição hierárquica do método da soma dos quadrados. Cada termo é ajustado apenas para o termo que o precede no modelo. O método da soma dos quadrados do Tipo I é normalmente utilizado para:

- Um modelo ANOVA balanceado, no qual quaisquer efeitos principais serão especificados antes de quaisquer efeitos de interação de primeira ordem, quaisquer efeitos de interação de primeira ordem são especificados antes de quaisquer efeitos de interação de segunda ordem, e assim por diante.
- Um modelo de regressão polinomial no qual quaisquer termos de ordem inferior são especificados antes de quaisquer termos de ordem superior.
- Um modelo puramente aninhado em que o primeiro efeito especificado é aninhado dentro do segundo efeito especificado, o segundo efeito especificado é aninhado dentro do terceiro, e assim por diante. (Esta forma de aninhamento pode ser especificada apenas utilizando a sintaxe).

**Tipo III.** O padrão. Este método calcula as somas dos quadrados de um efeito no design como as somas dos quadrados ajustados para quaisquer outros efeitos que não o contiver e é ortogonal para quaisquer efeitos (se houver) que o contiver. As somas dos quadrados do Tipo III possuem uma vantagem maior por serem invariáveis com relação às frequências de célula, desde que o formato geral de estimabilidade permaneça constante. Portanto, esse tipo 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 de médias dos quadrados ponderados de Yates. O método da soma dos quadrados do Tipo III é normalmente utilizado para:

- Quaisquer modelos listados no Tipo I.
- Quaisquer modelos balanceados ou não balanceados sem células vazias.

## Componentes de Variância Salvos no Novo Arquivo

É possível salvar alguns resultados deste procedimento em um novo arquivo de dados do IBM SPSS Estatísticas .

**Estimativas de componente de variância.** Salva as estimativas dos componentes de variância e estima os rótulos para um arquivo de dados ou conjunto de dados. Elas podem ser utilizadas no cálculo de mais estatísticas ou em análise adicional nos procedimentos GLM. Por exemplo, elas podem ser utilizadas para calcular intervalos de confiança ou hipóteses de teste.

**Covariância de componente.** Salva uma matriz de variância-covariância ou uma matriz de correlações em um arquivo de dados ou conjunto de dados. Disponível apenas se **Máxima verossimilhança** ou **Máxima verossimilhança restrita** tiver sido especificada.

**Destino para valores criados.** Permite especificar um nome do conjunto de dados ou um nome de arquivo externo para o arquivo que contém as estimativas e/ou a matriz de componente de variância. 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.

É possível utilizar o comando MATRIX para extrair os dados necessários a partir do arquivo de dados e, em seguida, calcular os intervalos de confiança ou executar testes.

## Recursos Adicionais do Comando VARCOMP

O idioma da sintaxe de comando também permite:

- Especificar efeitos aninhados no design (utilizando o subcomando DESIGN).
- Incluir valores omissos de usuário (utilizando o subcomando MISSING).
- Especificar critérios do ESP (utilizando o subcomando CRITERIA).

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

## Modelos mistos lineares

O procedimento de Modelos Lineares Mistos expande o modelo linear geral para que os dados possam exibir a variabilidade correlacionada e não constante. O modelo linear misto, portanto, fornece a flexibilidade de modelar não apenas as médias dos dados, mas as suas variâncias e covariâncias também.

O procedimento de Modelos Lineares Mistos também é uma ferramenta flexível para ajustar outros modelos que podem ser formulados como modelos lineares mistos. Esses modelos incluem modelos em vários níveis, modelos lineares hierárquicos e modelos de coeficiente aleatório.

### Exemplo:

Uma cadeia de mercearias está interessada nos efeitos de vários cupons sobre gastos dos clientes. Tomando uma amostra aleatória de seus clientes regulares, eles seguem o gasto de cada cliente durante 10 semanas. A cada semana, um cupom diferente é enviado pelo correio para os clientes. Modelos Lineares Mistos são usados para estimar o efeito de cupons diferentes no gasto, enquanto é

feito um ajuste para correlação em virtude de observações repetidas sobre cada assunto ao longo dos 10 semanas.

### Métodos

Estimação de máxima verossimilhança (ML) e de máxima verossimilhança restrita (REML).

### Estatísticas

Estatística descritiva: tamanhos de amostra, médias e desvios padrão da variável e de covariáveis dependentes para cada combinação de níveis distintos dos fatores. Informações de nível do fator: valores classificados dos níveis de cada fator e suas frequências. Além disso, estimativas de parâmetros e intervalos de confiança para efeitos fixos e testes de Wald e intervalos de confiança para parâmetros de matrizes de covariâncias. Somas dos quadrados do Tipo I e Tipo III podem ser usadas para avaliar hipóteses diferentes. O Tipo III é o padrão.

## Considerações sobre dados de Modelos lineares mistos

### Dados

A variável dependente deve ser quantitativa. Fatores devem ser categóricos e podem ter valores numéricos ou valores da sequência de caracteres. Covariáveis e a variável de ponderação devem ser quantitativas. Variáveis de assuntos e repetidas podem ser de qualquer tipo.

### Suposições

A variável dependente é considerada como linearmente relacionada aos fatores fixos, fatores aleatórios e covariáveis. Os efeitos fixos modelam a média da variável dependente. Os efeitos aleatórios modelam a estruturas de covariâncias da variável dependente. Efeitos aleatórios múltiplos são considerados independentes entre si e matrizes de covariâncias separadas serão calculadas para cada um; no entanto, termos modelos especificados no mesmo efeito aleatório podem ser correlacionados. As medidas repetidas modelam a estruturas de covariâncias dos resíduos. A variável dependente também é considerada como originada de uma distribuição normal.

### Procedimentos Relacionados

Use o procedimento Explorar para examinar os dados antes de executar uma análise. Se você não suspeitar que haja variabilidade correlacionada ou não constante, é possível usar o procedimento de GLM Univariado ou de Medidas Repetidas do GLM. É possível usar alternativamente o procedimento de Análise de Componentes de Variância se os efeitos aleatórios tiverem uma estruturas de covariâncias de componentes de variância e não houver medidas repetidas.

## Obtendo uma Análise de Modelos Lineares Mistos

1. Nos menus, escolha:

**Analisar > Modelos mistos > Linear...**

2. Opcionalmente, selecione uma ou mais variáveis de assunto.

3. Opcionalmente, selecione uma ou mais variáveis repetidas. Se as variáveis repetidas estiverem definidas, selecione um **Tipo de covariância repetida** da lista suspensa.

4. Opcionalmente, selecione uma ou mais variáveis de medida de Kronecker.

5. Opcionalmente, selecione uma estruturas de covariâncias de resíduo.

6. Clique em **Continuar**.

7. Selecione uma variável dependente.

8. Selecione pelo menos um fator ou covariável.

9. Clique em **Fixo** ou **Aleatório** e especifique pelo menos um modelo de efeitos fixos ou aleatórios.

Opcionalmente, selecione uma variável de ponderação.

## Modelos lineares mistos: Assuntos e Repetidos

Este diálogo permite selecionar variáveis que definem assuntos, observações repetidas, medidas de Kronecker e escolher uma estrutura de covariâncias para os resíduos.

## Sujeitos

Um assunto é uma unidade observacional que pode ser considerada independente de outros assuntos. Por exemplo, as leituras de pressão sanguínea de um paciente em um estudo médico podem ser consideradas independentes das leituras de outros pacientes. Definir assuntos torna-se particularmente importante quando existem medidas repetidas por assunto e você deseja modelar a correlação entre estas observações. Por exemplo, você pode esperar que as leituras de pressão sanguínea de um único paciente durante as visitas consecutivas para o médico sejam correlacionadas.

Assuntos também podem ser definidos pela combinação em nível de fator de variáveis múltiplas; por exemplo, é possível especificar *Sexo* e *Categoria de idade* como variáveis de assunto para modelar a crença de que *homens com idade superior a 65* são semelhantes entre si, mas independentes de *homens com menos de 65 e mulheres*.

Todas as variáveis especificadas na lista **Assuntos** são usadas para definir assuntos para a estrutura de covariâncias residual. É possível usar algumas ou todas as variáveis para definir assuntos para a estruturas de covariâncias de efeitos aleatórios.

## Repetido

As variáveis especificadas nessa lista são usadas para identificar observações repetidas. Por exemplo, uma única variável *Semana* pode identificar as 10 semanas de observações em um estudo médico ou *Mês* e *Dia* podem ser usados juntos para identificar observações diárias no curso de um ano.

## Tipo de covariância repetida

Isso especifica a estruturas de covariâncias para os resíduos. As estruturas disponíveis são como segue:

- Antedependência: Primeira ordem
- AR(1)
- Produto Direto AR1(UN\_AR1)
- Produto Direto não estruturado (UN\_UN)
- Simetria composta de produto direto (UN\_CS)
- AR(1): Heterogêneo
- ARMA (1,1)
- Simetria composta
- Simetria composta: Métrica de correlação
- Simetria composta: Heterogênea
- Diagonal
- Analítica de fator: Primeira ordem
- Analítica de fator: Primeira ordem, heterogênea
- Huynh-Feldt
- Identidade em escala
- Toeplitz
- Toeplitz: Heterogêneo
- Não estruturado
- Não estruturado: Métrica de correlação
- Espacial: Energia
- Espacial: Exponencial
- Espacial: Gaussiano
- Espacial: Linear
- Espacial: Log Linear
- Espacial: Esférico

### Medidas Kronecker

Selecione variáveis que especificam a estrutura de assunto para medidas de covariância de Kronecker e determinam como os erros de medida são correlacionados. O campo estará disponível somente quando um dos seguintes **Tipos de covariância repetida** for selecionado:

- Produto Direto AR1(UN\_AR1)
- Produto Direto não estruturado (UN\_UN)
- Simetria composta de produto direto (UN\_CS)

### Coordenadas de covariância espacial

As variáveis nesta lista especificam as coordenadas das observações repetidas quando um dos tipos de covariância espacial é selecionado para o tipo de covariância repetida.

Veja o tópico “Estruturas de Covariâncias” na página 102 para obter mais informações.

## Efeitos Fixos de Modelos Lineares Mistos

**Efeitos Fixos.** Não há nenhum modelo padrão; portanto, deve-se especificar explicitamente os efeitos fixos. Como alternativa, é possível criar termos aninhados e não aninhados.

**Incluir Intercepto.** O intercepto geralmente é incluído no modelo. Se for possível assumir a passagem de dados pela origem, é possível excluir o intercepto.

**Soma dos Quadrados.** O método de cálculo das somas dos quadrados. Para modelos sem nenhuma célula omissa, o método do Tipo III é mais comumente usado.

### Construir Termos Não Aninhados

Para os fatores e covariáveis selecionados:

**Fatorial.** Cria todas as interações e efeitos principais possíveis das variáveis selecionadas. Este é o padrão.

**Interação.** Cria o termo de interação de nível mais alto de todas as variáveis selecionadas.

**Efeitos principais.** Cria um termo dos principais efeitos para cada variável selecionada.

**Todos de 2 fatores.** Cria todas as interações de dois fatores das variáveis selecionadas.

**Todos de 3 fatores.** Cria todas as interações de três fatores possíveis das variáveis selecionadas.

**Todos de 4 fatores.** Cria todas as interações de quatro fatores possíveis das variáveis selecionadas.

**Todos de 5 fatores.** Cria todas as interações de cinco fatores possíveis das variáveis selecionadas.

### Construir Termos Aninhados

É possível construir termos aninhados para seu modelo neste procedimento. Os termos aninhados são úteis para modelar o efeito de um fator ou covariável cujos valores não interagem com os níveis de outro fator. Por exemplo, uma rede de supermercados pode seguir os hábitos de gastos de seus clientes em vários locais de loja. Como cada cliente frequenta apenas um desses locais, pode-se dizer que o efeito *Cliente* é **aninhado no** efeito *local da Loja*.

Além disso, é possível incluir efeitos de interações ou incluir diversos níveis de aninhamento no termo aninhado.

**Limitações.** Os termos aninhados possuem as restrições a seguir:

- Todos os fatores em uma interação devem ser exclusivos. Assim, se *A* for um fator, então especificar *A\*A* será inválido.
- Todos os fatores dentro de um efeito aninhado devem ser exclusivos. Assim, se *A* for um fator, então especificar *A(A)* será inválido.
- Nenhum efeito pode ser aninhado dentro de uma covariável. Dessa forma, se *A* for um fator e *X* for uma covariável, então, especificar *A(X)* é inválido.

## Soma dos Quadrados

Para o modelo, é possível escolher um tipo de somas dos quadrados. O Tipo III é o utilizado com mais frequência e é o padrão.

**Tipo I.** Esse método também é conhecido como decomposição hierárquica do método da soma dos quadrados. Cada termo é ajustado apenas para o termo que o precede no modelo. As somas dos quadrados do Tipo I são normalmente utilizadas para:

- Um modelo ANOVA balanceado, no qual quaisquer efeitos principais serão especificados antes de quaisquer efeitos de interação de primeira ordem, quaisquer efeitos de interação de primeira ordem são especificados antes de quaisquer efeitos de interação de segunda ordem, e assim por diante.
- Um modelo de regressão polinomial no qual quaisquer termos de ordem inferior são especificados antes de quaisquer termos de ordem superior.
- Um modelo puramente aninhado em que o primeiro efeito especificado é aninhado dentro do segundo efeito especificado, o segundo efeito especificado é aninhado dentro do terceiro, e assim por diante. (Esta forma de aninhamento pode ser especificada apenas utilizando a sintaxe).

**Tipo III.** O padrão. Este método calcula as somas dos quadrados de um efeito no design como as somas dos quadrados ajustados para quaisquer outros efeitos que não o contiver e é ortogonal para quaisquer efeitos (se houver) que o contiver. As somas dos quadrados do Tipo III possuem uma vantagem maior por serem invariáveis com relação às frequências de célula, desde 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 de médias dos quadrados ponderados de Yates. O método da soma dos quadrados do Tipo III é normalmente utilizado para:

- Quaisquer modelos listados no Tipo I.
- Quaisquer modelos balanceados ou não balanceados sem células vazias.

## Efeitos Aleatórios de Modelos Lineares Mistos

**Tipo de covariância.** Isso lhe permite especificar a estruturas de covariâncias para o modelo de efeitos aleatórios. Um matriz de covariâncias separada é estimada para cada efeito aleatório. As estruturas disponíveis são como segue:

- Antedependência: Primeira ordem
- AR(1)
- AR(1): Heterogêneo
- ARMA (1,1)
- Simetria composta
- Simetria composta: Métrica de correlação
- Simetria composta: Heterogênea
- Diagonal
- Analítica de fator: Primeira ordem
- Analítica de fator: Primeira ordem, heterogênea
- Huynh-Feldt
- Identidade em escala
- Toeplitz
- Toeplitz: Heterogêneo
- Não estruturado
- Não estruturado: Métrica de correlação
- Componentes de variância

Veja o tópico “Estruturas de Covariâncias” na página 102 para obter mais informações.

**Efeitos Aleatórios.** Não há nenhum modelo padrão; portanto, deve-se especificar explicitamente os efeitos aleatórios. Como alternativa, é possível criar termos aninhados e não aninhados. Também é possível escolher incluir um termo de interceptão no modelo de efeitos aleatórios.

É possível especificar modelos múltiplos de efeitos aleatórios. Após construir o primeiro modelo, clique em **Avançar** para construir o próximo modelo. Clique em **Anterior** para rolar de volta através de modelos existentes. Cada modelo de efeito aleatório é considerado como independente de qualquer outro modelo de efeito aleatório; isso é, matrizes de covariâncias separadas serão calculadas para cada um. Termos especificados no mesmo modelo de efeito aleatório podem ser correlacionados.

**Agrupamentos de Assunto.** As variáveis listadas são aquelas que você selecionou como variáveis de assunto na caixa de diálogo Selecionar Assuntos/Variáveis Repetidas. Escolha alguns ou todos esses para definir os assuntos para o modelo de efeitos aleatórios.

**Predições de parâmetro de exibição para este conjunto de efeitos aleatórios.** Especifica a exibição das estimativas de parâmetro de efeitos aleatórios.

## Estimação de Modelos Lineares

### Método

Selecione a estimativa de máxima verossimilhança ou de probabilidade máxima restrita.

### Graus de liberdade

Fornece opções para definir os graus de liberdade para todos os testes.

### Método residual

O método residual possui graus fixos de liberdade para todos os testes. Ele é útil se seu tamanho da amostra for suficientemente grande, ou os dados forem balanceados, ou se o modelo usar um tipo de covariância mais simples (por exemplo, identidade escalada ou diagonal).

### Aproximação Satterthwaite

O método Satterthwaite possui graus de liberdade de um campo nos testes. Ele é útil se seu tamanho da amostra for pequeno, ou os dados não forem balanceados ou se o modelo usar um tipo de covariância complicado (por exemplo, não estruturado).

### Aproximação Kenward-Roger

O método Kenward-Roger oferece um estimador de amostra pequeno mais preciso para a variância-covariância dos parâmetros de efeitos fixos e os graus de liberdade do denominador aproximados em testes t e testes F. O método introduz um fator de escala para a estatística F e estima-a e os graus de liberdade do denominador usando uma expansão de série Taylor para a estrutura aleatória estimada nos dados.

**Nota:** O método Kenward-Roger é usado na covariância baseada em modelo (em vez de covariância robusta). Quando o método Kenward-Roger e a covariância robusta são selecionados, o método Kenward-Roger é aplicado à covariância baseada em modelo e o seguinte aviso é apresentado: “Como o método Kenward-Roger foi selecionado, o método de covariância robusta foi mudado para método de covariância baseada em modelo”.

### Iterações

As opções a seguir estão disponíveis:

#### Máximo de iterações

Especifique um número inteiro não negativo.

#### Divisão máxima da etapa pela metade

A cada iteração, o tamanho do passo será reduzido por um fator de 0,5 até que o log da verossimilhança aumente ou o máximo de redução para metade do passo seja atingido. Especifique um número inteiro positivo.

#### Imprimir histórico de iteração para cada n etapa(s)

Exibe uma tabela contendo o valor da função de log da verossimilhança e as estimativas de parâmetros em cada  $n$  iteração, iniciando com a 0<sup>ésima</sup> iteração (as estimativas iniciais). Se optar



por imprimir o histórico de iteração, a última iteração será sempre impressa, independentemente do valor de  $n$ .

#### **Convergência de log da verossimilhança**

A convergência será assumida se a mudança absoluta ou mudança relativa na função de log da verossimilhança for menor do que o valor especificado, que deve ser não negativo. O critério não será utilizado se o valor especificado for igual a 0.

#### **Convergência de parâmetro**

A convergência será assumida se a mudança absoluta máxima ou a mudança relativa máxima nas estimativas de parâmetro for menor do que o valor especificado, que deve ser não negativo. O critério não será utilizado se o valor especificado for igual a 0.

#### **Convergência Hessiana**

Para uma especificação **Absoluta**, a convergência será assumida se uma estatística baseada na Hessiana for menor que o valor especificado. Para a especificação **Relativa**, a convergência será assumida se a estatística for menor que o produto do valor especificado e do valor absoluto do log da verossimilhança. O critério não será utilizado se o valor especificado for igual a 0.

#### **Máximo de etapas de escoragem**

Solicita utilizar o algoritmo de escoragem de Fisher até o número  $n$  de iterações. Especifique um número inteiro não negativo.

#### **Tolerância de singularidade**

Esse valor é usado como a tolerância na verificação de singularidade. Especifique um valor positivo.

## **Estatísticas de Modelos Lineares Mistos**

**Estatísticas de Sumarização.** Produz tabelas para:

- **Estatísticas descritivas.** Exibe os tamanhos da amostra, médias e desvios padrão da variável e covariáveis dependentes (se especificado). Essas estatísticas são exibidas para cada combinação de nível distinto dos fatores.
- **Sumarização de processamento de caso.** Exibe os valores classificados dos fatores, as variáveis de medida repetida, os assuntos de medida repetida e os assuntos de efeitos aleatórios e suas frequências.

**Estatísticas de Modelo.** Produz tabelas para:

- **Estimativas de parâmetro para efeitos fixos.** Exibe as estimativas de parâmetro de efeitos fixos e seus erros padrão aproximados.
- **Testes para parâmetros de covariância.** Exibe os erros padrão assintóticos e testes de Wald para os parâmetros de covariância.
- **Correlações de estimativas paramétrica.** Exibe a matriz de correlações assintóticas das estimativas paramétrica de efeitos fixos.
- **Covariâncias de estimativas paramétrica.** Exibe a matriz de covariância assintótica das estimativas paramétrica de efeitos fixos.
- **Covariâncias de efeitos aleatórios.** Exibe a matriz de covariâncias estimadas de efeitos aleatórios. Esta opção está disponível apenas quando pelo menos um efeito aleatório é especificado. Se uma variável de assunto for especificada para um efeito aleatório, então, o bloco comum será exibido.
- **Covariâncias de resíduos.** Exibe a matriz de covariância residual estimada. Esta opção está disponível apenas quando uma variável repetida tiver sido especificada. Se uma variável de assunto for especificada, o bloco comum será exibido.
- **Matriz de coeficiente de contraste.** Essa opção exibe as funções estimáveis usadas para testar os efeitos fixos e a hipótese customizada.

**Intervalo de confiança.** Esse valor é usado sempre que um intervalo de confiança é construído. Especifique um valor maior ou igual a 0 e menor que 100. O valor padrão é 95.

## Médias de EM de Modelos Lineares Mistos

**Médias Marginais Estimadas de Modelos Ajustados.** Esse grupo permite solicitar médias marginais estimadas previstas por modelo da variável dependente nas células e seus erros padrão para os fatores especificados. Além disso, é possível solicitar que níveis do fator dos principais efeitos sejam comparados.

- **Fator(es) e Interações entre Fatores.** Essa lista contém fatores e interações de fatores que foram especificados na caixa de diálogo Fixed (Fixo), além de um termo OVERALL. Termos modelo construídos a partir de covariáveis são excluídos dessa lista.
- **Exibir Médias para.** O procedimento calculará as médias marginais estimadas para fatores e interações entre fatores selecionadas na lista. Se a opção OVERALL for selecionada, as médias marginais estimadas da variável dependente serão exibidas, com o agrupamento de todos os fatores. Observe que qualquer fator selecionado ou interação entre fatores permanece selecionada a não ser que uma variável associada tenha sido removida da lista de Fatores na caixa de diálogo principal.
- **Comparar Efeitos Principais.** Essa opção permite solicitar comparação entre pares de níveis dos principais efeitos selecionados. O Ajustamento de Intervalo de Confiança permite aplicar um ajustamento nos intervalos de confiança e valores de significância para ter em conta as comparações múltiplas. Os métodos disponíveis são LSD (nenhum ajustamento), Bonferroni e Sidak. Finalmente, para cada fator, é possível selecionar uma categoria de referência com as quais comparações são feitas. Se nenhuma categoria de referência for selecionada, todas as comparações entre pares serão construídas. As opções para a categoria de referência são primeira, última ou customizada (nesse caso, você insere o valor da categoria de referência).

## Salvamento de Modelos Lineares Mistos

Essa caixa de diálogo permite salvar vários resultados de modelo no arquivo de trabalho.

**Valores Preditos Fixos.** Salva variáveis relacionadas com as médias de regressão sem os efeitos.

- **Valores preditos.** As médias de regressão sem os efeitos aleatórios.
- **Erros padrão.** Os erros padrão das estimativas.
- **Grau de liberdade.** Os graus de liberdade associados com as estimativas.

**Valores preditos & resíduos.** Salva variáveis relacionadas com o valor ajustado do modelo.

- **Valores preditos.** O valor ajustado do modelo.
- **Erros padrão.** Os erros padrão das estimativas.
- **Grau de liberdade.** Os graus de liberdade associados com as estimativas.
- **Residuais.** O valor dos dados menos o valor predito.

## Modelos Lineares Combinados-Exportação

Este diálogo permite exportar o conteúdo de tabelas de saída EBLUPs (Melhores Previsões Lineares Sem Viés) para conjuntos de dados ou arquivos .sav. O botão Exportar no diálogo Modelos Lineares Mistos será ativado se pelo menos um efeito aleatório for especificado por meio do diálogo Aleatório e a caixa "Exibir previsões de parâmetros para este bloco" for marcada.

### Exportar EBLUPS

Selecione um Destino: **Conjunto de dados** ou **Arquivo de dados**

Opcionalmente, forneça um nome.

Se diversos efeitos aleatórios com EBLUPs produzidos tiverem sido especificados, desmarque a caixa de seleção para que cada tabela de resultados apareça em um conjunto de dados ou arquivo separado

## Recursos Adicionais do Comando MIXED

O idioma da sintaxe de comando também permite:

- Especificar testes de efeitos versus uma combinação linear de efeitos ou um valor (utilizando o subcomando TEST).
- Incluir valores omissos de usuário (utilizando o subcomando MISSING).
- Calcular médias marginais estimadas para os valores especificados de covariáveis (utilizando a palavra-chave WITH do subcomando EMMEANS).
- Comparar efeitos principais simples de interações (utilizando o subcomando EMMEANS).

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

## Modelos lineares generalizados

O modelo linear generalizado expande o modelo linear geral de modo que a variável dependente seja linearmente relacionada com os fatores e as covariáveis através de uma função de ligação especificada. Além do mais, o modelo permite que a variável dependente tenha uma distribuição não normal. Ele cobre modelos estatísticos amplamente utilizados, tais como regressão linear para respostas distribuídas normalmente, modelos logísticos para dados binários, modelos log-lineares para dados de contagem, modelos log-log complementares para dados de sobrevivência com censura intervalar, além de muitos outros modelos estatísticos através de sua formulação de modelo muito geral.

**Exemplos.** Um companhia de navegação pode usar modelos lineares generalizados para ajustar uma regressão de Poisson às contagens de danos para diversos tipos de navios construídos em períodos de tempo diferentes e o modelo resultante pode ajudar a determinar quais tipos de navios são mais propensos a dano.

Uma empresa automobilística pode usar modelos lineares generalizados para ajustar uma regressão de gama às reclamações de danos para carros e o modelo resultante pode ajudar a determinar os fatores que contribuem mais com o tamanho da reclamação.

Pesquisadores médicos podem usar modelos lineares generalizados para ajustar uma regressão log-log complementar aos dados de sobrevivência censurados por intervalo para prever o tempo para recorrência para uma condição médica.

Considerações de Dados de Modelos Lineares Generalizados

**Dados.** A resposta pode ser escala, contagens, binária ou eventos em avaliações. Fatores são considerados categóricos. As covariáveis, a ponderação de escala e o offset são considerados escala.

**Suposições.** Casos são considerados observações independentes.

Para Obter um Modelo Linear Generalizado

Nos menus, escolha:

**Analisar > Modelos Lineares Generalizados > Modelos Lineares Generalizados...**

1. Especifique uma função de distribuição e de ligação (veja abaixo para obter detalhes sobre as várias opções).
2. Na guia Resposta, selecione uma variável dependente.
3. Na guia Preditores, selecione fatores e covariáveis para uso na predição da variável dependente.
4. Na guia Modelo, especifique efeitos do modelo usando os fatores e as covariáveis selecionados.

A guia Tipo de Modelo permite que você especifique a função de distribuição e de ligação para o seu modelo, fornecendo atalhos para vários modelos comuns que são categorizados por tipo de resposta.

Tipos de modelo

**Resposta Escalar.** As opções a seguir estão disponíveis:

- **Linear.** Especifica Normal como a distribuição e Identidade como a função de ligação.
- **Gama com ligação de log.** Especifica Gama como a distribuição e Log como a função de ligação.

**Resposta Ordinal.** As opções a seguir estão disponíveis:

- **Logística Ordinal.** Especifica Multinomial (ordinal) como a distribuição e Logit Acumulativo como a função de ligação.
- **Probita Ordinal.** Especifica Multinomial (ordinal) como a distribuição e Probita acumulativo como a função de ligação.

**Contagens.** As opções a seguir estão disponíveis:

- **Poisson log-linear.** Especifica Poisson como a distribuição e Log como a função de ligação.
- **Binomial negativa com ligação de log.** Especifica Binomial negativa (com um valor 1 para o parâmetro auxiliar) como a distribuição e Log como a função de ligação. Para que o procedimento estime o valor do parâmetro auxiliar, especifique um modelo customizado com distribuição Binomial negativa e selecione **Estimar valor** no Grupo paramétricas.

**Resposta Binária ou Dados de Eventos/Avaliações.** As opções a seguir estão disponíveis:

- **Logística binária.** Especifica Binomial como a distribuição e Logit como a função de ligação.
- **Probita binário.** Especifica Binomial como a distribuição e Probita como a função de ligação.
- **Sobrevivência censurada de intervalo.** Especifica Binomial como a distribuição e Log-log complementar como a função de ligação.

**Mistura.** As opções a seguir estão disponíveis:

- **Tweedie com ligação de log.** Especifica Tweedie como a distribuição e Log como a função de ligação.
- **Tweedie com ligação de identidade.** Especifica Tweedie como a distribuição e Identidade como a função de ligação.

**Customizado.** Especifique a sua própria combinação de distribuição e função de ligação.

#### Distribuição

Essa seleção especifica a distribuição da variável dependente. A capacidade para especificar uma distribuição não normal e a função de ligação de não identidade é a melhoria essencial do modelo linear generalizado sobre o modelo linear geral. Há muitas combinações de função de ligação de distribuição e diversas podem ser apropriadas para qualquer conjunto de dados especificado, de forma que a sua opção possa ser guiada por considerações teóricas a priori ou pela combinação que se ajuste melhor.

- **Binomial.** Essa distribuição é apropriada somente para variáveis que representam uma resposta binária ou um número de eventos.
- **Gama.** Essa distribuição é apropriada para variáveis com valores de escala positiva que são desviados em direção a valores positivos maiores. Se um valor de dados for menor que ou igual a 0 ou estiver omissa, então, o caso correspondente não será usado na análise.
- **Gaussiano Inversa.** Essa distribuição é apropriada para variáveis com valores de escala positiva que são desviados em direção a valores positivos maiores. Se um valor de dados for menor que ou igual a 0 ou estiver omissa, então, o caso correspondente não será usado na análise.
- **Binomial negativa.** Essa distribuição pode ser considerada como o número de avaliações necessárias para observar sucessos  $k$  e é apropriada para variáveis com valores de número inteiro não negativo. Se um valor dos dados for não inteiro, menor que 0 ou omissa, então, o caso correspondente não será usado na análise. O valor do parâmetro auxiliar da distribuição binomial negativa pode ser qualquer número maior que ou igual a 0; é possível configurá-lo para um valor fixo ou permitir que ele seja estimado pelo procedimento. Quando o parâmetro auxiliar é configurado como 0, usar essa distribuição é equivalente a usar a distribuição de Poisson.
- **Normal.** Isso é apropriado para variáveis de escala cujos valores usam uma distribuição simétrica, em forma de sino sobre um valor (média) central. A variável dependente deve ser numérica.
- **Poisson.** Essa distribuição pode ser considerada como o número de ocorrências de um evento de interesse em um período de tempo fixo e é apropriada para variáveis com valores de número inteiro não negativo. Se um valor dos dados for não inteiro, menor que 0 ou omissa, então, o caso correspondente não será usado na análise.
- **Tweedie.** Essa distribuição é apropriada para variáveis que podem ser representadas por misturas de Poisson de distribuições gama; a distribuição é "mista" no sentido de que ela combina propriedades

de distribuições contínuas (usa valores reais não negativos) e distribuições discretas (massa de probabilidade positiva a um valor único, 0). A variável dependente deve ser numérica, com valores dos dados maiores que ou iguais a zero. Se um valor de dados for menor que zero ou omissos, então, o caso correspondente não será usado na análise. O valor fixo do parâmetro da distribuição Tweedie pode ser qualquer número maior que um e menor que dois.

- **Multinomial.** Essa distribuição é apropriada para variáveis que representam uma resposta ordinal. A variável dependente pode ser numérica ou sequência de caracteres e ela deve ter pelo menos dois valores dos dados válidos distintos.

#### Funções de Ligação

A função de ligação é uma transformação da variável dependente que permite estimação do modelo. As funções a seguir estão disponíveis:

- **Identidade.**  $f(x)=x$ . A variável dependente não é transformada. Essa ligação pode ser usada com qualquer distribuição.
- **Log-log complementar.**  $f(x)=\log(-\log(1-x))$ . Isso é apropriado apenas com a distribuição binomial.
- **Cauchit Acumulativo.**  $f(x) = \tan(\pi (x - 0.5))$ , aplicado à probabilidade acumulativa de cada categoria da resposta. Isso é apropriado apenas com a distribuição multinomial.
- **Log-log complementar acumulativo.**  $f(x)=\ln(-\ln(1-x))$ , aplicado à probabilidade acumulativa de cada categoria da resposta. Isso é apropriado apenas com a distribuição multinomial.
- **Logit acumulativo.**  $f(x)=\ln(x / (1-x))$ , aplicado à probabilidade acumulativa de cada categoria da resposta. Isso é apropriado apenas com a distribuição multinomial.
- **Log-log negativo acumulativo.**  $f(x)=-\ln(-\ln(x))$ , aplicado à probabilidade acumulativa de cada categoria da resposta. Isso é apropriado apenas com a distribuição multinomial.
- **Probit Acumulativo.**  $f(x)=\Phi^{-1}(x)$ , aplicado à probabilidade acumulativa de cada categoria da resposta, em que  $\Phi^{-1}$  é a função de distribuição acumulativa normal padrão inversa. Isso é apropriado apenas com a distribuição multinomial.
- **Log.**  $f(x)=\log(x)$ . Essa ligação pode ser usada com qualquer distribuição.
- **Complemento de log.**  $f(x)=\log(1-x)$ . Isso é apropriado apenas com a distribuição binomial.
- **Logit.**  $f(x)=\log(x / (1-x))$ . Isso é apropriado apenas com a distribuição binomial.
- **Binomial negativa.**  $f(x)=\log(x / (x+k^{-1}))$ , em que  $k$  é o parâmetro auxiliar da distribuição binomial negativa. Isso é apropriado apenas com a distribuição binomial negativa.
- **Log-log negativo.**  $f(x)=-\log(-\log(x))$ . Isso é apropriado apenas com a distribuição binomial.
- **Poder das chances.**  $f(x)=[(x/(1-x))^{\alpha}-1]/\alpha$ , se  $\alpha \neq 0$ .  $f(x)=\log(x)$ , se  $\alpha=0$ .  $\alpha$  é a especificação numérica necessária e deve ser um número real. Isso é apropriado apenas com a distribuição binomial.
- **Probit.**  $f(x)=\Phi^{-1}(x)$ , em que  $\Phi^{-1}$  é a função de distribuição acumulativa normal padrão inversa. Isso é apropriado apenas com a distribuição binomial.
- **Poder.**  $f(x)=x^{\alpha}$ , se  $\alpha \neq 0$ .  $f(x)=\log(x)$ , se  $\alpha=0$ .  $\alpha$  é a especificação numérica necessária e deve ser um número real. Essa ligação pode ser usada com qualquer distribuição.

## Resposta de Modelos Lineares Generalizados

Em muitos casos, é possível simplesmente especificar uma variável dependente; no entanto, variáveis que tomam apenas dois valores e respostas que registram eventos em avaliações requerem atenção extra.

- **Resposta binária.** Quando a variável dependente toma apenas dois valores, é possível especificar a categoria de referência para estimação paramétrica. Uma variável de resposta binária pode ser sequência de caracteres ou numérica.
- **Número de eventos que ocorrem em um conjunto de avaliações.** Quando a resposta for um número de eventos que ocorrem em um conjunto de avaliações, a variável dependente conterá o número de eventos e será possível selecionar uma variável adicional contendo o número de avaliações. Como alternativa, se o número de avaliações for o mesmo em todos os sujeitos, então as avaliações poderão

ser especificadas utilizando um valor fixo. O número de avaliações deve ser maior ou igual ao número de eventos para cada caso. Eventos devem ser números inteiros não negativos e avaliações devem ser números inteiros positivos.

Para modelos multinomiais ordinais, é possível especificar a ordem de categoria da resposta: ascendente, decrescente ou dados (dados significa que o primeiro valor encontrado nos dados define a primeira categoria, o último valor exclusivo encontrado define a última categoria).

**Ponderação de Escala.** O parâmetro de escala é um parâmetro de modelo estimado relacionado à variância da resposta. As ponderações de escala são valores "conhecidos" que podem variar de observação para observação. Se a variável de ponderação de escala for especificada, o parâmetro de escala, que está relacionado à variância da resposta, será dividido por essa variável para cada observação. Casos com valores de ponderação de escala que forem menores ou iguais a 0 ou estiverem omissos não são usados na análise.

## Categoria de Referência de Modelos Lineares Generalizados

Para resposta binária, é possível escolher a categoria de referência para a variável dependente. Isso pode afetar uma determinada saída, como estimativas paramétricas e valores salvos, mas não deve alterar o ajuste do modelo. Por exemplo, se sua resposta binária utilizar os valores 0 e 1:

- Por padrão, o procedimento torna a última categoria (valor mais alto), ou 1, a categoria de referência. Nesta situação, as probabilidades de modelo salvas estimam a chance de um determinado caso assumir o valor 0, e as estimativas paramétrica devem ser interpretadas como relacionadas à probabilidade da categoria 0.
- Se você especificar a primeira categoria (valor mais baixo), ou 0, como a categoria de referência, então as probabilidades de modelo salvas estimarão a chance de um determinado caso assumir o valor 1.
- Se você especificar a categoria customizada e sua variável tiver rótulos definidos, será possível configurar a categoria de referência ao escolher um valor na lista. Isso poderá ser conveniente quando, no meio da especificação de um modelo, você não se lembrar exatamente como uma determinada variável foi codificada.

## Preditores de Modelos Lineares Generalizados

A guia Preditores permite que você especifique os fatores e as covariáveis usadas para construir efeitos do modelo e para especificar um offset opcional.

**Fatores.** Fatores são preditores categóricos; eles podem ser numéricos ou sequência de caracteres.

**Covariáveis.** Covariáveis são preditores de escala; eles devem ser numéricos.

*Nota:* Quando a resposta é binomial com formato binário, o procedimento calcula estatísticas de deviance e de qualidade de ajuste de qui-quadrado por subpopulações que são baseadas na classificação cruzada de valores observados dos fatores e covariáveis selecionados. É necessário manter o mesmo conjunto de preditores ao longo de várias execuções do procedimento para assegurar um número consistente de subpopulações.

**Offset.** O termo de offset é um preditor "estrutural". O seu coeficiente não é estimado pelo modelo, mas é assumido como tendo o valor 1; assim, os valores do offset são simplesmente incluídos no preditor linear da resposta. Isso é especialmente útil em modelos de regressão de Poisson, em que cada caso pode ter diferentes níveis de exposição ao evento de interesse.

Por exemplo, ao modelar as taxas de acidentes para motoristas individuais, há uma diferença importante entre um motorista que falhou em um acidente em três anos de experiência e um motorista que falhou em um acidente em 25 anos! O número de acidentes pode ser modelado como uma resposta de Poisson ou binomial negativa com uma ligação de log se o logaritmo natural da experiência do motorista estiver incluído como um termo de offset.

Outras combinações de distribuição e tipos de link iriam requerer outras transformações da variável de offset.

## Opções de Modelos Lineares Generalizados

Essas opções são aplicadas a todos os fatores especificados na guia Preditores.

**Valores omissos de usuário.** Os fatores devem ter valores válidos para um caso a serem incluídos na análise. Esses controles permitem que você decida se os valores omissos de usuário são tratados como válidos entre variáveis de fator.

**Ordem de categoria.** Isso é relevante para determinar o último nível de um fator, que pode ser associado com um parâmetro redundante no algoritmo de estimação. Mudar a ordem de categoria pode mudar os valores de efeitos de nível do fator, pois essas estimativas paramétrica são calculadas em relação ao "último" nível. Fatores podem ser ordenados em ordem ascendente a partir do valor mais baixo até o mais alto, em ordem decrescente a partir do valor mais alto até o mais baixo ou em "ordem de dados." Isso significa que o primeiro valor encontrado nos dados define a primeira categoria e o último valor exclusivo encontrado define a última categoria.

## Modelo de Modelos Lineares Generalizados

### Especificar efeitos do modelo

O modelo padrão é somente de interceptação; portanto, deve-se especificar explicitamente outros efeitos do modelo. Como alternativa, é possível criar termos aninhados e não aninhados.

### Termos não agrupados

Para os fatores e covariáveis selecionados:

#### Efeitos principais

Cria um termo dos principais efeitos para cada variável selecionada.

#### Interação

Cria o termo de interação de alto nível para todas as variáveis selecionadas.

#### Fatorial

Cria todas as interações e efeitos principais possíveis das variáveis selecionadas.

#### Todos de 2 fatores

Cria todas as interação de dois fatores das variáveis selecionadas.

#### Todos de 3 fatores

Cria todas as interações de três fatores possíveis das variáveis selecionadas.

#### Todos de 4 fatores

Cria todas as interações de quatro fatores possíveis das variáveis selecionadas.

#### Todos de 5 fatores

Cria todas as interações de cinco fatores possíveis das variáveis selecionadas.

### Termos agrupados

É possível construir termos aninhados para seu modelo neste procedimento. Os termos aninhados são úteis para modelar o efeito de um fator ou covariável cujos valores não interagem com os níveis de outro fator. Por exemplo, uma rede de supermercados pode seguir os hábitos de gastos de seus clientes em vários locais de loja. Como cada cliente frequenta somente um desses locais, o efeito do *Cliente* pode ser considerado **aninhado dentro** do efeito do *Local de armazenamento*.

Além disso, é possível incluir os efeitos de interação, como termos polinomiais envolvendo a mesma covariável, ou incluir diversos níveis de aninhamento no termo aninhado.

**Limitações:** Os termos aninhados possuem as restrições a seguir:

- Todos os fatores em uma interação devem ser exclusivos. Assim, se  $A$  for um fator, então especificar  $A^*A$  será inválido.
- Todos os fatores dentro de um efeito aninhado devem ser exclusivos. Assim, se  $A$  for um fator, então especificar  $A(A)$  será inválido.

- Nenhum efeito pode ser aninhado dentro de uma covariável. Dessa forma, se  $A$  for um fator e  $X$  for uma covariável, então, especificar  $A(X)$  é inválido.

## Estimação de Modelos Lineares Generalizados

**Estimação de Parâmetro.** Os controles nesse grupo permitem que você especifique método de estimação e forneça valores iniciais para as estimativas paramétrica.

- **Método.** É possível selecionar um método de estimação paramétrica. Escolha entre um método de Newton-Raphson, de escoragem de Fisher ou híbrido no qual iterações de escoragem de Fisher são executadas antes de alternar para o método de Newton-Raphson. Se a convergência for alcançada durante a fase de escoragem de Fisher do método híbrido antes de o número máximo de iterações Fisher ser alcançado, o algoritmo continua com o método Newton-Raphson.
- **Método paramétrica de escala.** É possível selecionar o método de estimação paramétrica de escala. A máxima verossimilhança conjuntamente estima o parâmetro de escala com os efeitos do modelo; observe que essa opção não é válida se a resposta tiver uma distribuição binomial negativa, de Poisson, binomial ou multinomial. As opções de deviance e de qui-quadrado de Pearson estimam o parâmetro de escala a partir do valor dessas estatísticas. Alternativamente, é possível especificar um valor fixo para o parâmetro de escala.
- **Valores iniciais.** O procedimento calculará automaticamente valores iniciais para parâmetros. Alternativamente, é possível especificar valores iniciais para as estimativas paramétrica.

**Matriz de covariâncias.** O estimador baseado em modelo é o negativo da inversa generalizada da matriz Hessiana. O estimador robusto (também chamado de Huber/White/sandwich) é um estimador baseado em modelo "corrigido" que fornece uma estimativa consistente da covariância, mesmo quando a especificação das funções de variância e de ligação estiver incorreta.

**Iterações.** As opções a seguir estão disponíveis:

- **Máximo de iterações.** O número máximo de iterações que o algoritmo executará. Especifique um número inteiro não negativo.
- **Máximo de redução para metade do passo.** A cada iteração, o tamanho do passo será reduzido por um fator de 0,5 até que o log da verossimilhança aumente ou o máximo de redução para metade do passo seja atingido. Especifique um número inteiro positivo.
- **Verificar a separação de pontos de dados.** Quando selecionado, o algoritmo executa testes para assegurar que as estimativas paramétrica têm valores exclusivos. A separação ocorre quando o procedimento pode produzir um modelo que ordena corretamente cada caso. Essa opção está disponível para respostas multinomiais e respostas binomiais com formato binário.

**Crítérios de Convergência.** As opções a seguir estão disponíveis

- **Convergência de Parâmetro.** Quando selecionada, o algoritmo é interrompido após uma iteração na qual uma mudança absoluta ou relativa nas estimativas paramétrica for menor que o valor especificado, que deve ser positivo.
- **Convergência de log da verossimilhança.** Quando selecionado, o algoritmo para após uma iteração na qual a mudança absoluta ou relativa na função de log da verossimilhança é inferior ao valor especificado, que deve ser positivo.
- **Convergência da Hessiana.** Para a Especificação absoluta, a convergência será assumida se uma estatística baseada na convergência da Hessiana for inferior ao valor positivo especificado. Para a Especificação relativa, a convergência será assumida se a estatística for inferior ao produto do valor positivo especificado e o valor absoluto do log da verossimilhança.

**Tolerância à singularidade.** Matrizes singulares (ou não inversíveis) possuem colunas linearmente dependentes que podem causar sérios problemas para o algoritmo de estimação. Como até mesmo matrizes quase singulares podem levar a resultados ruins, o procedimento tratará como singular uma matriz cuja determinante for menor que a tolerância. Especifique um valor positivo.



## Valores de Modelos Lineares Generalizados

Se valores iniciais forem especificados, eles devem ser fornecidos para todos os parâmetros (incluindo parâmetros redundantes) no modelo. No conjunto de dados, a ordenação de variáveis da esquerda para a direita deve ser: *RowType\_*, *VarName\_*, *P1*, *P2*, ..., em que *RowType\_* e *VarName\_* são variáveis de sequência de caracteres e *P1*, *P2*, ... são variáveis numéricas que correspondem a uma lista ordenada dos parâmetros.

- Valores iniciais são fornecidos em um registro com valor *EST* para variável *RowType\_*; os valores iniciais reais são fornecidos sob variáveis *P1*, *P2*, .... O procedimento ignora todos os registros para os quais *RowType\_* tem um valor diferente de *EST* bem como qualquer registro além da primeira ocorrência de *RowType\_* igual a *EST*.
- O intercepto, se incluído no modelo, ou parâmetros de limite, se a resposta tiver uma distribuição multinomial, devem ser os primeiros valores iniciais listados.
- O parâmetro de escala *e*, se a resposta tiver uma distribuição binomial negativa, o parâmetro binomial negativo, devem ser os últimos valores iniciais especificados.
- Se Arquivo Dividido estiver em vigor, então as variáveis devem começar com a variável ou variáveis de divisão de arquivo na ordem especificada ao criar o Arquivo Dividido, seguido por *RowType\_*, *VarName\_*, *P1*, *P2*, ... como acima. Divisões devem ocorrer no conjunto de dados especificado na mesma ordem que no conjunto de dados original.

*Nota:* Os nomes de variáveis *P1*, *P2*, ... não são necessários; o procedimento aceitará qualquer nome de variável válido para os parâmetros porque o mapeamento de variáveis para parâmetros é baseado na posição da variável, não o nome da variável. Qualquer variável além do último parâmetro será ignorada.

A estrutura do arquivo para os valores iniciais é a mesma que aquela usada ao exportar o modelo como dados; assim, é possível usar os valores finais a partir de uma execução do procedimento como entrada em uma execução subsequente.

## Estatísticas de Modelos Lineares Generalizados

**Efeitos do Modelo.** As opções a seguir estão disponíveis:

- **Tipo de análise.** Especifique o tipo de análise a ser produzido. A análise de tipo I é geralmente apropriada quando você tem razões a priori para ordenar preditores no modelo, enquanto o Tipo III é geralmente mais aplicável. Estatísticas Wald ou de razão de verossimilhança são calculadas com base na seleção no grupo de Estatísticas Qui-Quadrado.
- **Intervalos de confiança.** Especifique um nível de confiança maior que 50 e menor que 100. Intervalos Wald são baseados na suposição de que parâmetros têm uma distribuição normal assintótica; intervalos de probabilidade de perfil são mais exatos, mas podem ser dispendiosos computacionalmente. O nível de tolerância para intervalos de probabilidade de perfil são os critérios usados para parar o algoritmo iterativo usado para calcular os intervalos.
- **Função de log da verossimilhança.** Isso controla o formato de exibição da função de log da verossimilhança. A função integral inclui um termo adicional que é constante com relação às estimativas de parâmetros; ela não tem nenhum efeito em estimativa paramétrica e é deixada fora da exibição em alguns produtos de software.

**Imprimir.** O resultado a seguir está disponível:

- **Sumarização do processamento de caso.** Exibe o número e a porcentagem de casos incluídos e excluídos da análise e a tabela de Sumarização de Dados Correlacionados.
- **Estatísticas descritivas.** Exibe estatísticas descritivas e informações de resumo sobre a variável dependente, as covariáveis e fatores.
- **Informações de modelo.** Exibe o nome do conjunto de dados, variável dependente ou eventos e variáveis de avaliação, variável de offset, variável de ponderação de escala, distribuição de probabilidade e função de ligação.

- **Qualidade de estatísticas de ajuste.** Exibe deviance e deviance escalado, qui-quadrado de Pearson e qui-quadrado de Pearson escalado, log da verossimilhança, critério de informações de Akaike (AIC), AIC corrigido de amostra finita (AICC), critério de informações bayesiano (BIC) e AIC consistente (CAIC).
- **Estatísticas de sumarização do modelo.** Exibe testes de ajuste do modelo, incluindo estatísticas de razão de verossimilhança para o teste de omnibus de ajuste do modelo e estatísticas para os contrastes Tipo I ou III para cada efeito.
- **Estimativas paramétrica.** Exibe estimativas de parâmetro, estatísticas do teste correspondentes e intervalos de confiança. É possível opcionalmente exibir estimativas de parâmetros exponenciado além das estimativas de parâmetros bruto.
- **Matriz de covariâncias para estimativas de parâmetros.** Exibe a matriz de covariâncias paramétrica estimado.
- **Matriz de correlações para estimativas de parâmetros.** Exibe a matriz de correlações paramétrica estimado.
- **Matrizes de coeficiente de contraste (L).** Exibe coeficientes de contraste para os efeitos padrão e para as médias marginais estimadas, se solicitados na guia Médias de EM.
- **Funções estimáveis gerais.** Exibe as matrizes para gerar as matrizes de coeficiente de contraste (L).
- **Histórico de iteração.** Exibe o histórico de iteração para as estimativas de parâmetros e o log da verossimilhança e imprime a última avaliação do vetor gradiente e a matriz Hessiana. A tabela de históricos de iteração exibe estimativas de parâmetros para cada  $n^{\text{a}}$  iteração que inicia com a 0<sup>a</sup> iteração (as estimativas iniciais), em que  $n$  é o valor do intervalo de impressão. Se o histórico de iteração for solicitado, então a última iteração será sempre exibida, independentemente de  $n$ .
- **Teste de multiplicadores de Lagrange de parâmetro de escala ou de parâmetro auxiliar binomial negativo.** Exibe estatísticas do teste de multiplicadores de Lagrange para avaliar a validade de um parâmetro de escala que é calculado usando o deviance ou qui-quadrado de Pearson ou configurado em um número fixo, para as distribuições normais, gama, Gaussianas inversas e de Tweedie. Para a distribuição binomial negativa, isso testa o parâmetro auxiliar fixo.

## Médias de EM de Modelos Lineares Generalizados

Essa guia permite que você exiba as médias marginais estimadas para níveis de fatores e interações entre fatores. Também é possível solicitar que a média geral estimada seja exibida. Médias marginais estimadas não estão disponíveis para modelos multinomiais ordinais.

**Fatores e Interações.** Essa lista contém fatores especificados na guia Preditores e interações entre fatores especificadas na guia Modelo. Covariáveis são excluídas dessa lista. Termos podem ser selecionados diretamente a partir dessa lista ou combinados em um termo de interação usando o botão **Por \***.

**Exibir Médias Para.** Médias estimadas são calculadas para os fatores e as interações entre fatores selecionados. O contraste determina como testes de hipótese são configurados para comparar as médias estimadas. O contraste simples requer uma categoria de referência ou um nível do fator com relação aos quais os outros são comparados.

- **Pairwise.** Comparações entre pares são calculadas para combinações de todos os níveis dos fatores especificados ou implícitos. Esse é o único contraste disponível para interações entre fatores.
- **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.
- **Desvio.** Cada nível do fator é comparado com a média global. Contrastes de desvio não são ortogonais.
- **Diferença.** Compara a média de cada nível (exceto do primeiro) com a média dos níveis anteriores. Às vezes são chamados de contraste de Helmert reversos.
- **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 utilizados para estimar as tendências polinomiais.

**Escala.** Médias marginais estimadas podem ser calculadas para a resposta, com base na escala original da variável dependente ou para o preditor linear, com base na variável dependente conforme transformada pela função de ligação.

**Ajuste para Várias Comparações.** Ao executar testes de hipótese com vários contrastes, o nível de significância global pode ser ajustado a partir dos níveis de significância para os contrastes incluídos. Esse grupo permite que você escolha o método de ajuste.

- **Diferença menos significativa.** Esse método não controla a probabilidade geral de rejeitar a hipótese de que algum contraste linear seja diferente dos valores de hipótese nula.
- *Bonferroni*. Este método ajusta o nível de significância observada para o fato de que diversos contrastes estão sendo testados.
- *Bonferroni Sequencial*. Este é um procedimento de Bonferroni de rejeição sequencialmente decrescente que tende ser muito menos conservador em termos de rejeição de hipóteses individuais, mas mantém o mesmo nível de significância geral.
- *Sidak*. Este método fornece limites mais apertados que a abordagem de Bonferroni.
- *Sidak Sequencial*. Este é um procedimento de Sidak de rejeição sequencialmente decrescente que tende ser muito menos conservador em termos de rejeição de hipóteses individuais, mas mantém o mesmo nível de significância geral.

## Salvamento de Modelos Lineares Generalizados

Itens verificados são salvos com o nome especificado; é possível escolher sobrescrever variáveis existentes com o mesmo nome que as novas variáveis ou evitar conflitos de nome por sufixos de apêndice para tornar os nomes das novas variáveis exclusivos.

### Valor previsto da média de resposta

Salva valores preditos pelo modelo para cada caso na métrica de resposta original. Quando a distribuição de resposta for binomial e a variável dependente for binária, o procedimento salva probabilidades preditas. Quando a distribuição de resposta for multinomial, o rótulo do item torna-se **Probabilidade predita acumulativa** e o procedimento salva a probabilidade predita acumulativa para cada categoria da resposta, exceto a última, até o número de categorias especificadas para salvar.

### Limite inferior do intervalo de confiança para a média de resposta

Salva o limite inferior do intervalo de confiança para a média da resposta. Quando a distribuição de resposta for multinomial, o rótulo do item torna-se **Limite inferior de intervalo de confiança para probabilidade predita acumulativa** e o procedimento salva o limite inferior para cada categoria da resposta, exceto a última, até o número de categorias especificadas para salvar.

### Limite superior do intervalo de confiança para a média de resposta

Salva o limite superior do intervalo de confiança para a média da resposta. Quando a distribuição de resposta for multinomial, o rótulo do item torna-se **Limite superior de intervalo de confiança para probabilidade predita acumulativa** e o procedimento salva o limite superior para cada categoria da resposta, exceto a última, até o número de categorias especificadas para salvar.

### Categoria predita

Para modelos com distribuição binomial e variável dependente binária ou distribuição multinomial, isso salva a categoria de resposta predita para cada caso. Essa opção não está disponível para outras distribuições de resposta.

### Valor predito de preditor linear

Salva valores preditos pelo modelo para cada caso na métrica do preditor linear (resposta transformada através da função de ligação especificada). Quando a distribuição de resposta for multinomial, o procedimento salva o valor predito para cada categoria da resposta, exceto a última, até o número de categorias especificadas para salvar.

**Erro padrão estimado de valor predito de preditor linear**

Quando a distribuição de resposta for multinomial, o procedimento salva o erro padrão estimado para cada categoria da resposta, exceto a última, até o número de categorias especificadas para salvar.

Os itens a seguir não estão disponíveis quando a distribuição de respostas é multinomial.

**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.

**Valor de ponto 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  $(N-1)/N$ .

**Resíduo**

A diferença entre um valor observado e o valor predito pelo modelo

**Resíduo de Pearson**

A raiz quadrada da contribuição de um caso para a estatística qui-quadrado de Pearson, com o sinal do resíduo bruto.

**Resíduo de Pearson padronizado**

O resíduo de Pearson multiplicado pela raiz quadrada do inverso do produto do parâmetro de escala e alavancagem  $h$  para o caso.

**Resíduos de desvio**

A raiz quadrada da contribuição de um caso para a estatística de Deviance, com o sinal do resíduo bruto.

**Resíduos de desvio padronizado**

O resíduo de Deviance multiplicado pela raiz quadrada do inverso do produto do parâmetro de escala e alavancagem  $h$  para o caso.

**Resíduo de probabilidade**

A raiz quadrada de uma média ponderada (baseada na alavancagem do caso) dos quadrados dos resíduos de Pearson padronizados e de Deviance padronizados, com o sinal do resíduo bruto.

## Exportação de Modelos Lineares Generalizados

**Exportar modelo como dados.** Grava um conjunto de dados em formato do IBM SPSS Estatísticas contendo a correlação paramétrica ou matriz de covariâncias com estimativas paramétrica, erros padrão, valores de significância e graus de liberdade. A ordem de variáveis no arquivo de matriz é como segue.

- **Variáveis de divisão.** Se usada, qualquer variável definindo divisões.
- **RowType\_.** Leva valores (e rótulos de valor) *COV* (covariâncias), *CORR* (correlações), *EST* (estimativas de parâmetros), *SE* (erros padrão), *SIG* (níveis de significância) e *DF* (graus de liberdade de design de amostragem). Há um caso separado com tipo de linha *COV* (ou *CORR*) para cada parâmetro de modelo, mais um caso separado para cada um dos outros tipos de linha.
- **VarName\_.** Obtém valores *P1*, *P2*, ..., correspondente a uma lista ordenada de todos os parâmetros estimados do modelo (exceto a escala ou os parâmetros binomiais negativos), para tipos de linha *COV* ou *CORR*, com rótulos de valores correspondentes às cadeias de parâmetros mostradas na tabela Parameter estimates (Estimativas de parâmetros). As células são em branco para outros tipos de linha.
- **P1, P2, ...** Essas variáveis correspondem a uma lista ordenada de todos os parâmetros de modelo (incluindo a escala e parâmetros binomiais negativos, conforme apropriado), com rótulos de variável correspondentes às sequências de caracteres paramétrica mostradas na tabela de Estimativas paramétrica e usam valores de acordo com o tipo de linha.

Para parâmetros redundantes, todas as covariâncias são configuradas como zero, as correlações são configuradas para o valor omissivo do sistema; todas as estimativas paramétrica são configuradas em zero; e todos os erros padrão, níveis de significância e graus de liberdade de resíduo são configurados para o valor omissivo do sistema.

Para o parâmetro de escala, as covariâncias, correlações, o nível de significância e graus de liberdade são configurados para o valor omisso do sistema. Se o parâmetro de escala for estimado através de máxima verossimilhança, o erro padrão é especificado; caso contrário, ele é configurado para o valor omisso do sistema.

Para o parâmetro binomial negativo, as covariâncias, correlações, o nível de significância e graus de liberdade são configurados para o valor omisso do sistema. Se o parâmetro binomial negativo for estimado através de máxima verossimilhança, o erro padrão é especificado; caso contrário, ele é configurado para o valor omisso do sistema.

Se houver divisões, então, a lista paramétrica deve ser acumulada através de todas as divisões. Em uma determinada divisão, alguns parâmetros podem ser irrelevantes; isso não é o mesmo que redundante. Para parâmetros irrelevantes, todas as covariâncias ou correlações, estimativas paramétrica, erros padrão, níveis de significância e graus de liberdade são configurados para o valor omisso do sistema.

É possível usar esse arquivo de matriz como os valores iniciais para estimação de modelo adicional; observe que esse arquivo não é imediatamente utilizável para análises adicionais em outros procedimentos que leem um arquivo de matriz, a menos que esses procedimentos aceitem todos os tipos de linha exportados aqui. Mesmo assim, você deve tomar cuidado para que todos os parâmetros nesse arquivo de matriz tenham o mesmo significado para o procedimento que estiver lendo o arquivo.

**Exportar modelo como XML.** Salva as estimativas paramétrica e a matriz de covariância paramétrica, se selecionado, 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.

## Variáveis Adicionais de Comando do GENLIN

O idioma da sintaxe de comando também permite:

- Especificar valores iniciais para estimativas paramétrica como uma lista de números (usando o subcomando CRITERIA).
- Fixar covariáveis em valores diferentes de suas médias ao calcular médias marginais estimadas (usando o subcomando EMMEANS).
- Especificar contrastes polinomiais customizados para médias marginais estimadas (usando o subcomando EMMEANS).
- Especificar um subconjunto dos fatores para os quais médias marginais estimadas são exibidas para serem comparadas usando o tipo de contraste especificado (usando as palavras-chave TABLES e COMPARE do subcomando EMMEANS).

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

## Equações de estimativa generalizada

O procedimento de Equações de Estimativa Generalizada amplia o modelo linear generalizado para permitir a análise de medidas repetidas ou outras observações correlacionadas, como dados em cluster.

**Exemplo,** Funcionários da saúde pública podem usar equações de estimativa generalizada para ajustar uma regressão logística de medida repetida para estudar efeitos da poluição atmosférica em crianças.

Considerações de Dados de Equações de Estimativa Generalizada

**Dados.** A resposta pode ser escala, contagens, binária ou eventos em avaliações. Fatores são considerados categóricos. As covariáveis, a ponderação de escala e o offset são considerados escala. Variáveis usadas para definir assuntos ou medidas repetidas dentre-sujeitos não podem ser usadas para definir a resposta, mas podem servir outras funções no modelo.

**Suposições.** Casos são considerados dependentes dentre-sujeitos e independentes entre sujeitos. A matriz de correlações que representa as dependências dentre-sujeitos é estimada como parte do modelo.

Obtendo Equações de Estimativa Generalizada

Nos menus, escolha:

**Analisar > Modelos Lineares Generalizados > Equações de Estimativa Generalizada...**

1. Selecione uma ou mais variáveis de assunto (veja abaixo para obter opções adicionais).

A combinação de valores das variáveis especificadas deve definir exclusivamente **sujeitos** dentro do conjunto de dados. Por exemplo, uma única variável *ID do Paciente* deve ser suficiente para definir sujeitos em um único hospital, mas a combinação de *ID do Hospital* e *ID do Paciente* pode ser necessária se os números de identificação do paciente não forem exclusivos entre os hospitais. Em uma configuração de medidas repetidas, várias observações são registradas para cada sujeito, de modo que cada sujeito possa ocupar diversos casos no conjunto de dados.

2. Na guia Tipo de Modelo, especifique uma função de distribuição e de ligação.
3. Na guia Resposta, selecione uma variável dependente.
4. Na guia Preditores, selecione fatores e covariáveis para uso na predição da variável dependente.
5. Na guia Modelo, especifique efeitos do modelo usando os fatores e as covariáveis selecionados.

Opcionalmente, na guia Repetida é possível especificar:

**Variáveis dentre-sujeitos.** A combinação de valores das variáveis dentre-sujeitos define a ordenação das medições dentre sujeitos; assim, a combinação de variáveis dentre sujeitos e de sujeito define exclusivamente cada medição. Por exemplo, a combinação de *Período*, *ID do Hospital* e *ID do Paciente* define, para cada caso, uma consulta particular para um paciente específico em um hospital específico.

Se o conjunto de dados já estiver classificado de modo que as medições repetidas de cada sujeito ocorra em um bloco de casos contíguos e na ordem adequada, não é estritamente necessário especificar uma variável dentre sujeitos e é possível cancelar a seleção de **Classificar casos por variáveis de assunto e dentre sujeitos** e salvar o tempo de processamento necessário para executar a classificação (provisória). Geralmente, é uma boa ideia fazer uso de variáveis dentre-sujeitos para assegurar a ordenação adequada de medições.

Variáveis de sujeito e dentre-sujeitos não podem ser usadas para definir a resposta, mas podem executar outras funções no modelo. Por exemplo, *ID do Hospital* poderia ser usado como um fator no modelo.

**Matriz de Covariâncias.** O estimador baseado em modelo é o negativo da inversa generalizada da matriz Hessiana. O estimador robusto (também chamado de estimador Huber/White/sandwich) é um estimador baseado em modelo "corrigido" que fornece uma estimativa consistente da covariância, mesmo quando a matriz de correlação de trabalho estiver incorreta. Essa especificação aplica-se aos parâmetros na parte de modelo linear das equações de estimativa generalizada, enquanto a especificação na guia Estimação aplica-se somente ao modelo linear generalizado inicial.

**Matriz de Correlação de Trabalho.** Essa matriz de correlações representa as dependências dentre-sujeitos. Seu tamanho é determinado pelo número de medições e, assim, pela combinação de valores de variáveis dentre-sujeitos. É possível especificar uma das estruturas a seguir:

- **Independente.** Medidas repetidas são não correlacionadas.
- **AR(1).** Medidas repetidas têm um relacionamento autorregressivo de primeira ordem. A correlação entre quaisquer dois elementos é igual a  $\rho$  para elementos adjacentes,  $\rho^2$  para elementos que são separados por um terceiro, e assim por diante. é restringido de forma que  $-1 < \rho < 1$ .
- **Intercambiável.** Essa estrutura tem correlações homogêneas entre elementos. Ela também é conhecida como estrutura de simetria composta.
- **Dependente de M.** Medições consecutivas têm um coeficiente de correlação comum, pares de medições separados por um terceiro têm um coeficiente de correlação comum, e assim por diante, por meio de pares de medições separados por outras medições  $m-1$ . Por exemplo, se você der aos alunos testes padronizados a cada ano do 3º até o 7º grau. Essa estrutura supõe que os escores do 3º e 4º, 4º e 5º, 5º e 6º e 6º e 7º graus terão a mesma correlação; 3º e 5º, 4º e 6º e 5º e 7º terão a mesma correlação; 3º e 6º e 4º e 7º terão a mesma correlação. Medições com separação maior que  $m$  são consideradas não correlacionadas. Ao escolher essa estrutura, especifique um valor  $m$  inferior à ordem da matriz de correlação de trabalho.
- **Não estruturada.** Essa é uma matriz de correlações completamente geral.

Por padrão, o procedimento ajustará as estimativas de correlação pelo número paramétricas não redundantes. Remover esse ajuste pode ser desejável se você deseja que as estimativas sejam invariáveis para mudanças de replicação de nível de sujeito nos dados.

- **Máximo de iterações.** O número máximo de iterações que o algoritmo de equações de estimativa generalizada executará. Especifique um número inteiro não negativo. Essa especificação aplica-se aos parâmetros na parte de modelo linear das equações de estimativa generalizada, enquanto a especificação na guia [Estimação](#) aplica-se somente ao modelo linear generalizado inicial.
- **Atualizar matriz.** Elementos na matriz de correlação de trabalho são estimados com base nas estimativas paramétrica, que são atualizadas em cada iteração do algoritmo. Se a matriz de correlação de trabalho não for atualizada de modo algum, a matriz de correlação de trabalho inicial será usada por todo o processo de estimação. Se a matriz for atualizada, é possível especificar o intervalo de iteração no qual atualizar elementos da matriz de correlação de trabalho. Especificar um valor maior que 1 pode reduzir tempo de processamento.

**Crítérios de convergência.** Essas especificações aplicam-se aos parâmetros na parte de modelo linear das equações de estimativa generalizada, enquanto a especificação na guia [Estimação](#) aplica-se somente ao modelo linear generalizado inicial.

- **Convergência de Parâmetro.** Quando selecionada, o algoritmo é interrompido após uma iteração na qual uma mudança absoluta ou relativa nas estimativas paramétrica for menor que o valor especificado, que deve ser positivo.
- **Convergência da Hessiana.** A convergência será assumida se uma estatística com base na Hessiana for menor que o valor especificado, que deve ser positivo.

## Tipo de Modelo de Equações de Estimativa Generalizada

A guia Tipo de Modelo permite que você especifique a distribuição e a função de ligação para o seu modelo, fornecendo atalhos para vários modelos comuns que são categorizados por tipo de resposta.

Tipos de modelo

**Resposta Escalar.** As opções a seguir estão disponíveis:

- **Linear.** Especifica Normal como a distribuição e Identidade como a função de ligação.
- **Gama com ligação de log.** Especifica Gama como a distribuição e Log como a função de ligação.

**Resposta Ordinal.** As opções a seguir estão disponíveis:

- **Logística Ordinal.** Especifica Multinomial (ordinal) como a distribuição e Logit Acumulativo como a função de ligação.
- **Probita Ordinal.** Especifica Multinomial (ordinal) como a distribuição e Probita acumulativo como a função de ligação.

**Contagens.** As opções a seguir estão disponíveis:

- **Poisson log-linear.** Especifica Poisson como a distribuição e Log como a função de ligação.
- **Binomial negativa com ligação de log.** Especifica Binomial negativa (com um valor 1 para o parâmetro auxiliar) como a distribuição e Log como a função de ligação. Para que o procedimento estime o valor do parâmetro auxiliar, especifique um modelo customizado com distribuição Binomial negativa e selecione **Estimar valor** no Grupo paramétricas.

**Resposta Binária ou Dados de Eventos/Avaliações.** As opções a seguir estão disponíveis:

- **Logística binária.** Especifica Binomial como a distribuição e Logit como a função de ligação.
- **Probita binário.** Especifica Binomial como a distribuição e Probita como a função de ligação.
- **Sobrevivência censurada de intervalo.** Especifica Binomial como a distribuição e Log-log complementar como a função de ligação.

**Mistura.** As opções a seguir estão disponíveis:

- **Tweedie com ligação de log.** Especifica Tweedie como a distribuição e Log como a função de ligação.

- **Tweedie com ligação de identidade.** Especifica Tweedie como a distribuição e Identidade como a função de ligação.

**Customizado.** Especifique a sua própria combinação de distribuição e função de ligação.

#### Distribuição

Essa seleção especifica a distribuição da variável dependente. A capacidade para especificar uma distribuição não normal e a função de ligação de não identidade é a melhoria essencial do modelo linear generalizado sobre o modelo linear geral. Há muitas combinações de função de ligação de distribuição e diversas podem ser apropriadas para qualquer conjunto de dados especificado, de forma que a sua opção possa ser guiada por considerações teóricas a priori ou pela combinação que se ajuste melhor.

- **Binomial.** Essa distribuição é apropriada somente para variáveis que representam uma resposta binária ou um número de eventos.
- **Gama.** Essa distribuição é apropriada para variáveis com valores de escala positiva que são desviados em direção a valores positivos maiores. Se um valor de dados for menor que ou igual a 0 ou estiver omissa, então, o caso correspondente não será usado na análise.
- **Gaussiano Inversa.** Essa distribuição é apropriada para variáveis com valores de escala positiva que são desviados em direção a valores positivos maiores. Se um valor de dados for menor que ou igual a 0 ou estiver omissa, então, o caso correspondente não será usado na análise.
- **Binomial negativa.** Essa distribuição pode ser considerada como o número de avaliações necessárias para observar sucessos  $k$  e é apropriada para variáveis com valores de número inteiro não negativo. Se um valor dos dados for não inteiro, menor que 0 ou omissa, então, o caso correspondente não será usado na análise. O valor do parâmetro auxiliar da distribuição binomial negativa pode ser qualquer número maior que ou igual a 0; é possível configurá-lo para um valor fixo ou permitir que ele seja estimado pelo procedimento. Quando o parâmetro auxiliar é configurado como 0, usar essa distribuição é equivalente a usar a distribuição de Poisson.
- **Normal.** Isso é apropriado para variáveis de escala cujos valores usam uma distribuição simétrica, em forma de sino sobre um valor (média) central. A variável dependente deve ser numérica.
- **Poisson.** Essa distribuição pode ser considerada como o número de ocorrências de um evento de interesse em um período de tempo fixo e é apropriada para variáveis com valores de número inteiro não negativo. Se um valor dos dados for não inteiro, menor que 0 ou omissa, então, o caso correspondente não será usado na análise.
- **Tweedie.** Essa distribuição é apropriada para variáveis que podem ser representadas por misturas de Poisson de distribuições gama; a distribuição é "mista" no sentido de que ela combina propriedades de distribuições contínuas (usa valores reais não negativos) e distribuições discretas (massa de probabilidade positiva a um valor único, 0). A variável dependente deve ser numérica, com valores dos dados maiores que ou iguais a zero. Se um valor de dados for menor que zero ou omissa, então, o caso correspondente não será usado na análise. O valor fixo do parâmetro da distribuição Tweedie pode ser qualquer número maior que um e menor que dois.
- **Multinomial.** Essa distribuição é apropriada para variáveis que representam uma resposta ordinal. A variável dependente pode ser numérica ou sequência de caracteres e ela deve ter pelo menos dois valores dos dados válidos distintos.

#### Função de ligação

A função de ligação é uma transformação da variável dependente que permite estimação do modelo. As funções a seguir estão disponíveis:

- **Identidade.**  $f(x)=x$ . A variável dependente não é transformada. Essa ligação pode ser usada com qualquer distribuição.
- **Log-log complementar.**  $f(x)=\log(-\log(1-x))$ . Isso é apropriado apenas com a distribuição binomial.
- **Cauchit Acumulativo.**  $f(x) = \tan(\pi (x - 0.5))$ , aplicado à probabilidade acumulativa de cada categoria da resposta. Isso é apropriado apenas com a distribuição multinomial.
- **Log-log complementar acumulativo.**  $f(x)=\ln(-\ln(1-x))$ , aplicado à probabilidade acumulativa de cada categoria da resposta. Isso é apropriado apenas com a distribuição multinomial.



- **Logit acumulativo.**  $f(x)=\ln(x / (1-x))$ , aplicado à probabilidade acumulativa de cada categoria da resposta. Isso é apropriado apenas com a distribuição multinomial.
- **Log-log negativo acumulativo.**  $f(x)=-\ln(-\ln(x))$ , aplicado à probabilidade acumulativa de cada categoria da resposta. Isso é apropriado apenas com a distribuição multinomial.
- **Probit Acumulativo.**  $f(x)=\Phi^{-1}(x)$ , aplicado à probabilidade acumulativa de cada categoria da resposta, em que  $\Phi^{-1}$  é a função de distribuição acumulativa normal padrão inversa. Isso é apropriado apenas com a distribuição multinomial.
- **Log.**  $f(x)=\log(x)$ . Essa ligação pode ser usada com qualquer distribuição.
- **Complemento de log.**  $f(x)=\log(1-x)$ . Isso é apropriado apenas com a distribuição binomial.
- **Logit.**  $f(x)=\log(x / (1-x))$ . Isso é apropriado apenas com a distribuição binomial.
- **Binomial negativa.**  $f(x)=\log(x / (x+k^{-1}))$ , em que  $k$  é o parâmetro auxiliar da distribuição binomial negativa. Isso é apropriado apenas com a distribuição binomial negativa.
- **Log-log negativo.**  $f(x)=-\log(-\log(x))$ . Isso é apropriado apenas com a distribuição binomial.
- **Poder das chances.**  $f(x)=[(x/(1-x))^{\alpha}-1]/\alpha$ , se  $\alpha \neq 0$ .  $f(x)=\log(x)$ , se  $\alpha=0$ .  $\alpha$  é a especificação numérica necessária e deve ser um número real. Isso é apropriado apenas com a distribuição binomial.
- **Probit.**  $f(x)=\Phi^{-1}(x)$ , em que  $\Phi^{-1}$  é a função de distribuição acumulativa normal padrão inversa. Isso é apropriado apenas com a distribuição binomial.
- **Poder.**  $f(x)=x^{\alpha}$ , se  $\alpha \neq 0$ .  $f(x)=\log(x)$ , se  $\alpha=0$ .  $\alpha$  é a especificação numérica necessária e deve ser um número real. Essa ligação pode ser usada com qualquer distribuição.

## Resposta de Equações de Estimativa Generalizada

Em muitos casos, é possível simplesmente especificar uma variável dependente; no entanto, variáveis que tomam apenas dois valores e respostas que registram eventos em avaliações requerem atenção extra.

- **Resposta binária.** Quando a variável dependente toma apenas dois valores, é possível especificar a categoria de referência para estimação paramétrica. Uma variável de resposta binária pode ser sequência de caracteres ou numérica.
- **Número de eventos que ocorrem em um conjunto de avaliações.** Quando a resposta for um número de eventos que ocorrem em um conjunto de avaliações, a variável dependente conterá o número de eventos e será possível selecionar uma variável adicional contendo o número de avaliações. Como alternativa, se o número de avaliações for o mesmo em todos os sujeitos, então as avaliações poderão ser especificadas utilizando um valor fixo. O número de avaliações deve ser maior ou igual ao número de eventos para cada caso. Eventos devem ser números inteiros não negativos e avaliações devem ser números inteiros positivos.

Para modelos multinomiais ordinais, é possível especificar a ordem de categoria da resposta: ascendente, decrescente ou dados (dados significa que o primeiro valor encontrado nos dados define a primeira categoria, o último valor exclusivo encontrado define a última categoria).

**Ponderação de Escala.** O parâmetro de escala é um parâmetro de modelo estimado relacionado à variância da resposta. As ponderações de escala são valores "conhecidos" que podem variar de observação para observação. Se a variável de ponderação de escala for especificada, o parâmetro de escala, que está relacionado à variância da resposta, será dividido por essa variável para cada observação. Casos com valores de ponderação de escala que forem menores ou iguais a 0 ou estiverem omissos não são usados na análise.

## Categoria de Referência de Equações de Estimativa Generalizada

Para resposta binária, é possível escolher a categoria de referência para a variável dependente. Isso pode afetar uma determinada saída, como estimativas paramétricas e valores salvos, mas não deve alterar o ajuste do modelo. Por exemplo, se sua resposta binária utilizar os valores 0 e 1:

- Por padrão, o procedimento torna a última categoria (valor mais alto), ou 1, a categoria de referência. Nesta situação, as probabilidades de modelo salvas estimam a chance de um determinado caso

assumir o valor 0, e as estimativas paramétrica devem ser interpretadas como relacionadas à probabilidade da categoria 0.

- Se você especificar a primeira categoria (valor mais baixo), ou 0, como a categoria de referência, então as probabilidades de modelo salvas estimarão a chance de um determinado caso assumir o valor 1.
- Se você especificar a categoria customizada e sua variável tiver rótulos definidos, será possível configurar a categoria de referência ao escolher um valor na lista. Isso poderá ser conveniente quando, no meio da especificação de um modelo, você não se lembrar exatamente como uma determinada variável foi codificada.

## Preditores de Equações de Estimativa Generalizada

A guia Preditores permite que você especifique os fatores e as covariáveis usadas para construir efeitos do modelo e para especificar um offset opcional.

**Fatores.** Fatores são preditores categóricos; eles podem ser numéricos ou sequência de caracteres.

**Covariáveis.** Covariáveis são preditores de escala; eles devem ser numéricos.

*Nota:* Quando a resposta é binomial com formato binário, o procedimento calcula estatísticas de deviance e de qualidade de ajuste de qui-quadrado por subpopulações que são baseadas na classificação cruzada de valores observados dos fatores e covariáveis selecionados. É necessário manter o mesmo conjunto de preditores ao longo de várias execuções do procedimento para assegurar um número consistente de subpopulações.

**Offset.** O termo de offset é um preditor "estrutural". O seu coeficiente não é estimado pelo modelo, mas é assumido como tendo o valor 1; assim, os valores do offset são simplesmente incluídos no preditor linear da resposta. Isso é especialmente útil em modelos de regressão de Poisson, em que cada caso pode ter diferentes níveis de exposição ao evento de interesse.

Por exemplo, ao modelar as taxas de acidentes para motoristas individuais, há uma diferença importante entre um motorista que falhou em um acidente em três anos de experiência e um motorista que falhou em um acidente em 25 anos! O número de acidentes pode ser modelado como uma resposta de Poisson ou binomial negativa com uma ligação de log se o logaritmo natural da experiência do motorista estiver incluído como um termo de offset.

Outras combinações de distribuição e tipos de link iriam requerer outras transformações da variável de offset.

## Opções de Equações de Estimativa Generalizada

Essas opções são aplicadas a todos os fatores especificados na guia Preditores.

**Valores omissos de usuário.** Os fatores devem ter valores válidos para um caso a serem incluídos na análise. Esses controles permitem que você decida se os valores omissos de usuário são tratados como válidos entre variáveis de fator.

**Ordem de categoria.** Isso é relevante para determinar o último nível de um fator, que pode ser associado com um parâmetro redundante no algoritmo de estimação. Mudar a ordem de categoria pode mudar os valores de efeitos de nível do fator, pois essas estimativas paramétrica são calculadas em relação ao "último" nível. Fatores podem ser ordenados em ordem ascendente a partir do valor mais baixo até o mais alto, em ordem decrescente a partir do valor mais alto até o mais baixo ou em "ordem de dados." Isso significa que o primeiro valor encontrado nos dados define a primeira categoria e o último valor exclusivo encontrado define a última categoria.

## Modelo de Equações de Estimativa Generalizada

**Especificar efeitos do modelo.** O modelo padrão é somente de interceptação; portanto, deve-se especificar explicitamente outros efeitos do modelo. Como alternativa, é possível criar termos aninhados e não aninhados.

Termos não aninhados

Para os fatores e covariáveis selecionados:

**Efeitos principais.** Cria um termo dos principais efeitos para cada variável selecionada.

**Interação.** Cria o termo de interação de alto nível para todas as variáveis selecionadas.

**Fatorial.** Cria todas as interações e efeitos principais possíveis das variáveis selecionadas.

**Todos os 2 fatores.** Cria todas as interações de dois fatores das variáveis selecionadas.

**Todos 3 fatores.** Cria todas as interações de três fatores possíveis das variáveis selecionadas.

**Todos os 4 fatores.** Cria todas as interações de quatro fatores possíveis das variáveis selecionadas.

**Todos os 5 fatores.** Cria todas as interações de cinco fatores possíveis das variáveis selecionadas.

Termos aninhados

É possível construir termos aninhados para seu modelo neste procedimento. Os termos aninhados são úteis para modelar o efeito de um fator ou covariável cujos valores não interagem com os níveis de outro fator. Por exemplo, uma rede de supermercados pode seguir os hábitos de gastos de seus clientes em vários locais de loja. Como cada cliente frequenta somente um desses locais, o efeito do *Cliente* pode ser considerado **aninhado dentro** do efeito do *Local de armazenamento*.

Além disso, é possível incluir efeitos de interações ou incluir diversos níveis de aninhamento no termo aninhado.

**Limitações.** Os termos aninhados possuem as restrições a seguir:

- Todos os fatores em uma interação devem ser exclusivos. Assim, se  $A$  for um fator, então especificar  $A*A$  será inválido.
- Todos os fatores dentro de um efeito aninhado devem ser exclusivos. Assim, se  $A$  for um fator, então especificar  $A(A)$  será inválido.
- Nenhum efeito pode ser aninhado dentro de uma covariável. Dessa forma, se  $A$  for um fator e  $X$  for uma covariável, então, especificar  $A(X)$  é inválido.

**Intercepto.** O intercepto geralmente é incluído no modelo. Se for possível assumir a passagem de dados pela origem, é possível excluir o intercepto.

Modelos com a distribuição ordinal de multinômio não têm um termo de intercepção único; em vez disso, existem parâmetros de limite que definem pontos de transição entre categorias adjacentes. Os limites são sempre incluídos no modelo.

## Estimação de Equações de Estimativa Generalizada

**Estimação de Parâmetro.** Os controles nesse grupo permitem que você especifique método de estimação e forneça valores iniciais para as estimativas paramétrica.

- **Método.** É possível selecionar um método de estimação paramétrica; escolha entre Newton-Raphson, escoragem de Fisher ou um método híbrido no qual as iterações da escoragem de Fisher são executadas antes da comutação para o método de Newton-Raphson. Se a convergência for alcançada durante a fase de escoragem de Fisher do método híbrido antes de o número máximo de iterações Fisher ser alcançado, o algoritmo continua com o método Newton-Raphson.
- **Método de Parâmetro de Escala.** É possível selecionar o método de estimação paramétrica de escala.

A máxima verossimilhança conjuntamente estima o parâmetro de escala com os efeitos do modelo; observe que essa não é válida se a resposta tiver uma distribuição binomial negativa, de Poisson ou binomial. Como o conceito de probabilidade não entra em equações de estimativa generalizada, essa especificação aplica-se apenas ao modelo linear generalizado inicial; essa estimativa paramétrica de escala é, então, passada para as equações de estimativa generalizada, que atualizam o parâmetro de escala pelo qui-quadrado de Pearson dividido por seus graus de liberdade.

As opções de deviance e de qui-quadrado de Pearson estimam o parâmetro de escala a partir do valor dessas estatísticas no modelo linear generalizado inicial; essa estimativa paramétrica de escala é, então, passada para as equações de estimativa generalizada, que o tratam como fixo.

Como alternativa, especifique um valor fixo para o parâmetro da escala. Ele será tratado como fixo ao estimar o modelo linear generalizado inicial e as equações de estimativa generalizada.

- **Valores iniciais.** O procedimento calculará automaticamente valores iniciais para parâmetros. Alternativamente, é possível especificar valores iniciais para as estimativas paramétrica.

As iterações e os critérios de convergência especificados nessa guia são aplicáveis somente ao modelo linear generalizado inicial. Para critérios de estimação usados no ajuste das equações de estimativa generalizada, veja a guia Repetida.

**Iterações.** As opções a seguir estão disponíveis:

- **Máximo de iterações.** O número máximo de iterações que o algoritmo executará. Especifique um número inteiro não negativo.
- **Máximo de redução para metade do passo.** A cada iteração, o tamanho do passo será reduzido por um fator de 0,5 até que o log da verossimilhança aumente ou o máximo de redução para metade do passo seja atingido. Especifique um número inteiro positivo.
- **Verificar a separação de pontos de dados.** Quando selecionado, o algoritmo executa testes para assegurar que as estimativas paramétrica têm valores exclusivos. A separação ocorre quando o procedimento pode produzir um modelo que ordena corretamente cada caso. Essa opção está disponível para respostas multinomiais e respostas binomiais com formato binário.

**Críticos de Convergência.** As opções a seguir estão disponíveis

- **Convergência de Parâmetro.** Quando selecionada, o algoritmo é interrompido após uma iteração na qual uma mudança absoluta ou relativa nas estimativas paramétrica for menor que o valor especificado, que deve ser positivo.
- **Convergência de log da verossimilhança.** Quando selecionado, o algoritmo para após uma iteração na qual a mudança absoluta ou relativa na função de log da verossimilhança é inferior ao valor especificado, que deve ser positivo.
- **Convergência da Hessiana.** Para a Especificação absoluta, a convergência será assumida se uma estatística baseada na convergência da Hessiana for inferior ao valor positivo especificado. Para a Especificação relativa, a convergência será assumida se a estatística for inferior ao produto do valor positivo especificado e o valor absoluto do log da verossimilhança.

**Tolerância à singularidade.** Matrizes singulares (ou não inversíveis) possuem colunas linearmente dependentes que podem causar sérios problemas para o algoritmo de estimação. Como até mesmo matrizes quase singulares podem levar a resultados ruins, o procedimento tratará como singular uma matriz cuja determinante for menor que a tolerância. Especifique um valor positivo.

## Valores Iniciais de Equações de Estimativa Generalizada

O procedimento estima um modelo linear generalizado inicial e as estimativas a partir desse modelo são usadas como valores iniciais para as estimativas paramétrica na parte do modelo linear das equações de estimativa generalizada. Valores iniciais não são necessários para a matriz de correlação de trabalho porque elementos de matriz são baseados nas estimativas paramétrica. Valores iniciais especificados nessa caixa de diálogo são usados como o ponto de início para o modelo linear generalizado inicial, não as equações de estimativa generalizada, a menos que as Iterações máximas na guia Estimação estejam configuradas para 0.

Se valores iniciais forem especificados, eles devem ser fornecidos para todos os parâmetros (incluindo parâmetros redundantes) no modelo. No conjunto de dados, a ordenação de variáveis da esquerda para a direita deve ser: *RowType\_*, *VarName\_*, *P1*, *P2*, ..., em que *RowType\_* e *VarName\_* são variáveis de sequência de caracteres e *P1*, *P2*, ... são variáveis numéricas que correspondem a uma lista ordenada dos parâmetros.

- Valores iniciais são fornecidos em um registro com valor *EST* para variável *RowType\_*; os valores iniciais reais são fornecidos sob variáveis *P1*, *P2*, .... O procedimento ignora todos os registros para os quais *RowType\_* tem um valor diferente de *EST* bem como qualquer registro além da primeira ocorrência de *RowType\_* igual a *EST*.

- O intercepto, se incluído no modelo, ou parâmetros de limite, se a resposta tiver uma distribuição multinomial, devem ser os primeiros valores iniciais listados.
- O parâmetro de escala e, se a resposta tiver uma distribuição binomial negativa, o parâmetro binomial negativo, devem ser os últimos valores iniciais especificados.
- Se Arquivo Dividido estiver em vigor, então as variáveis devem começar com a variável ou variáveis de divisão de arquivo na ordem especificada ao criar o Arquivo Dividido, seguido por *RowType\_*, *VarName\_*, *P1*, *P2*, ... como acima. Divisões devem ocorrer no conjunto de dados especificado na mesma ordem que no conjunto de dados original.

*Nota:* Os nomes de variáveis *P1*, *P2*, ... não são necessários; o procedimento aceitará qualquer nome de variável válido para os parâmetros porque o mapeamento de variáveis para parâmetros é baseado na posição da variável, não o nome da variável. Qualquer variável além do último parâmetro será ignorada.

A estrutura do arquivo para os valores iniciais é a mesma que aquela usada ao exportar o modelo como dados; assim, é possível usar os valores finais a partir de uma execução do procedimento como entrada em uma execução subsequente.

## Estatísticas de Equações de Estimativa Generalizada

**Efeitos do Modelo.** As opções a seguir estão disponíveis:

- **Tipo de análise.** Especifique o tipo de análise a ser produzido para efeitos do modelo de teste. A análise de tipo I é geralmente apropriada quando você tem razões a priori para ordenar preditores no modelo, enquanto o Tipo III é geralmente mais aplicável. Estatísticas Wald ou de escore generalizado são calculadas com base na seleção feita no grupo Estatísticas Qui-quadrado.
- **Intervalos de confiança.** Especifique um nível de confiança maior que 50 e menor que 100. Intervalos Wald são sempre produzidos independentemente do tipo de estatística qui-quadrado selecionado e são baseados na suposição de que parâmetros têm uma distribuição normal assintótica.
- **Função quasi-verossimilhança de log.** Isso controla o formato de exibição da função quasi-verossimilhança de log. A função integral inclui um termo adicional que é constante com relação às estimativas de parâmetros; ela não tem nenhum efeito em estimativa paramétrica e é deixada fora da exibição em alguns produtos de software.

**Imprimir.** A saída a seguir está disponível.

- **Sumarização do processamento de caso.** Exibe o número e a porcentagem de casos incluídos e excluídos da análise e a tabela de Sumarização de Dados Correlacionados.
- **Estatísticas descritivas.** Exibe estatísticas descritivas e informações de resumo sobre a variável dependente, as covariáveis e fatores.
- **Informações de modelo.** Exibe o nome do conjunto de dados, variável dependente ou eventos e variáveis de avaliação, variável de offset, variável de ponderação de escala, distribuição de probabilidade e função de ligação.
- **Qualidade de estatísticas de ajuste.** Exibe duas extensões de Critério de Informações de Akaike para seleção de modelo: Quasi-verossimilhança sob o critério de modelo de independência (QIC) para escolher a melhor estrutura de correlação e uma outra medida de QIC para escolher o melhor subconjunto de preditores.
- **Estatísticas de sumarização do modelo.** Exibe testes de ajuste do modelo, incluindo estatísticas de razão de verossimilhança para o teste de omnibus de ajuste do modelo e estatísticas para os contrastes Tipo I ou III para cada efeito.
- **Estimativas paramétrica.** Exibe estimativas de parâmetro, estatísticas do teste correspondentes e intervalos de confiança. É possível opcionalmente exibir estimativas de parâmetros exponenciado além das estimativas de parâmetros bruto.
- **Matriz de covariâncias para estimativas de parâmetros.** Exibe a matriz de covariâncias paramétrica estimado.
- **Matriz de correlações para estimativas de parâmetros.** Exibe a matriz de correlações paramétrica estimado.

- **Matrizes de coeficiente de contraste (L).** Exibe coeficientes de contraste para os efeitos padrão e para as médias marginais estimadas, se solicitados na guia Médias de EM.
- **Funções estimáveis gerais.** Exibe as matrizes para gerar as matrizes de coeficiente de contraste (L).
- **Histórico de iteração.** Exibe o histórico de iteração para as estimativas de parâmetros e o log da verossimilhança e imprime a última avaliação do vetor gradiente e a matriz Hessiana. A tabela de históricos de iteração exibe estimativas de parâmetros para cada  $n^{\text{a}}$  iteração que inicia com a 0<sup>a</sup> iteração (as estimativas iniciais), em que  $n$  é o valor do intervalo de impressão. Se o histórico de iteração for solicitado, então a última iteração será sempre exibida, independentemente de  $n$ .
- **Matriz de correlação de trabalho.** Exibe os valores da matriz que representa as dependências dentre-sujeitos. A sua estrutura depende das especificações na guia Repetida.

## Médias de EM de Equações de Estimativa Generalizada

Essa guia permite que você exiba as médias marginais estimadas para níveis de fatores e interações entre fatores. Também é possível solicitar que a média geral estimada seja exibida. Médias marginais estimadas não estão disponíveis para modelos multinomiais ordinais.

**Fatores e Interações.** Essa lista contém fatores especificados na guia Preditores e interações entre fatores especificadas na guia Modelo. Covariáveis são excluídas dessa lista. Termos podem ser selecionados diretamente a partir dessa lista ou combinados em um termo de interação usando o botão **Por \***.

**Exibir Médias Para.** Médias estimadas são calculadas para os fatores e as interações entre fatores selecionados. O contraste determina como testes de hipótese são configurados para comparar as médias estimadas. O contraste simples requer uma categoria de referência ou um nível do fator com relação aos quais os outros são comparados.

- **Pairwise.** Comparações entre pares são calculadas para combinações de todos os níveis dos fatores especificados ou implícitos. Esse é o único contraste disponível para interações entre fatores.
- **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.
- **Desvio.** Cada nível do fator é comparado com a média global. Contrastes de desvio não são ortogonais.
- **Diferença.** Compara a média de cada nível (exceto do primeiro) com a média dos níveis anteriores. Às vezes são chamados de contraste de Helmert reversos.
- **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 utilizados para estimar as tendências polinomiais.

**Escala.** Médias marginais estimadas podem ser calculadas para a resposta, com base na escala original da variável dependente ou para o preditor linear, com base na variável dependente conforme transformada pela função de ligação.

**Ajuste para Várias Comparações.** Ao executar testes de hipótese com vários contrastes, o nível de significância global pode ser ajustado a partir dos níveis de significância para os contrastes incluídos. Esse grupo permite que você escolha o método de ajuste.

- **Diferença menos significativa.** Esse método não controla a probabilidade geral de rejeitar a hipótese de que algum contraste linear seja diferente dos valores de hipótese nula.
- **Bonferroni.** Este método ajusta o nível de significância observada para o fato de que diversos contrastes estão sendo testados.
- **Bonferroni Sequencial.** Este é um procedimento de Bonferroni de rejeição sequencialmente decrescente que tende ser muito menos conservador em termos de rejeição de hipóteses individuais, mas mantém o mesmo nível de significância geral.

- *Sidak*. Este método fornece limites mais apertados que a abordagem de Bonferroni.
- *Sidak Sequencial*. Este é um procedimento de Sidak de rejeição sequencialmente decrescente que tende ser muito menos conservador em termos de rejeição de hipóteses individuais, mas mantém o mesmo nível de significância geral.

## Salvamento de Equações de Estimativa Generalizada

Itens verificados são salvos com o nome especificado; é possível escolher sobrescrever variáveis existentes com o mesmo nome que as novas variáveis ou evitar conflitos de nome por sufixos de apêndice para tornar os nomes das novas variáveis exclusivos.

- **Valor predito de média de resposta.** Salva valores preditos pelo modelo para cada caso na métrica de resposta original. Quando a distribuição de resposta for binomial e a variável dependente for binária, o procedimento salva probabilidades preditas. Quando a distribuição de resposta for multinomial, o rótulo do item torna-se **Probabilidade predita acumulativa** e o procedimento salva a probabilidade predita acumulativa para cada categoria da resposta, exceto a última, até o número de categorias especificadas para salvar.
- **Limite inferior de intervalo de confiança para média de resposta.** Salva o limite inferior do intervalo de confiança para a média da resposta. Quando a distribuição de resposta for multinomial, o rótulo do item torna-se **Limite inferior de intervalo de confiança para probabilidade predita acumulativa** e o procedimento salva o limite inferior para cada categoria da resposta, exceto a última, até o número de categorias especificadas para salvar.
- **Limite superior de intervalo de confiança para média de resposta.** Salva o limite superior do intervalo de confiança para a média da resposta. Quando a distribuição de resposta for multinomial, o rótulo do item torna-se **Limite superior de intervalo de confiança para probabilidade predita acumulativa** e o procedimento salva o limite superior para cada categoria da resposta, exceto a última, até o número de categorias especificadas para salvar.
- **Categoria predita.** Para modelos com distribuição binomial e variável dependente binária ou distribuição multinomial, isso salva a categoria de resposta predita para cada caso. Essa opção não está disponível para outras distribuições de resposta.
- **Valor predito de preditor linear.** Salva valores preditos pelo modelo para cada caso na métrica do preditor linear (resposta transformada através da função de ligação especificada). Quando a distribuição de resposta for multinomial, o procedimento salva o valor predito para cada categoria da resposta, exceto a última, até o número de categorias especificadas para salvar.
- **Erro padrão estimado de valor predito de preditor linear.** Quando a distribuição de resposta for multinomial, o procedimento salva o erro padrão estimado para cada categoria da resposta, exceto a última, até o número de categorias especificadas para salvar.

Os itens a seguir não estão disponíveis quando a distribuição de respostas é multinomial.

- *Resíduo bruto*. A diferença entre um valor observado e o valor predito pelo modelo
- **Resíduo de Pearson**. A raiz quadrada da contribuição de um caso para a estatística qui-quadrado de Pearson, com o sinal do resíduo bruto.

## Exportação de Equações de Estimativa Generalizada

**Exportar modelo como dados.** Grava um conjunto de dados em formato do IBM SPSS Estatísticas contendo a correlação paramétrica ou matriz de covariâncias com estimativas paramétrica, erros padrão, valores de significância e graus de liberdade. A ordem de variáveis no arquivo de matriz é como segue.

- **Variáveis de divisão.** Se usada, qualquer variável definindo divisões.
- **RowType\_.** Leva valores (e rótulos de valor) *COV* (covariâncias), *CORR* (correlações), *EST* (estimativas de parâmetros), *SE* (erros padrão), *SIG* (níveis de significância) e *DF* (graus de liberdade de design de amostragem). Há um caso separado com tipo de linha *COV* (ou *CORR*) para cada parâmetro de modelo, mais um caso separado para cada um dos outros tipos de linha.
- **VarName\_.** Obtém valores *P1*, *P2*, ..., correspondente a uma lista ordenada de todos os parâmetros estimados do modelo (exceto a escala ou os parâmetros binomiais negativos), para tipos de linha

COV ou CORR, com rótulos de valores correspondentes às cadeias de parâmetros mostradas na tabela Parameter estimates (Estimativas de parâmetros). As células são em branco para outros tipos de linha.

- **P1, P2, ...** Essas variáveis correspondem a uma lista ordenada de todos os parâmetros de modelo (incluindo a escala e parâmetros binomiais negativos, conforme apropriado), com rótulos de variável correspondentes às sequências de caracteres paramétrica mostradas na tabela de Estimativas paramétrica e usam valores de acordo com o tipo de linha.

Para parâmetros redundantes, todas as covariâncias são configuradas como zero, as correlações são configuradas para o valor omissivo do sistema; todas as estimativas paramétrica são configuradas em zero; e todos os erros padrão, níveis de significância e graus de liberdade de resíduo são configurados para o valor omissivo do sistema.

Para o parâmetro de escala, as covariâncias, correlações, o nível de significância e graus de liberdade são configurados para o valor omissivo do sistema. Se o parâmetro de escala for estimado através de máxima verossimilhança, o erro padrão é especificado; caso contrário, ele é configurado para o valor omissivo do sistema.

Para o parâmetro binomial negativo, as covariâncias, correlações, o nível de significância e graus de liberdade são configurados para o valor omissivo do sistema. Se o parâmetro binomial negativo for estimado através de máxima verossimilhança, o erro padrão é especificado; caso contrário, ele é configurado para o valor omissivo do sistema.

Se houver divisões, então, a lista paramétricas deve ser acumulada através de todas as divisões. Em uma determinada divisão, alguns parâmetros podem ser irrelevantes; isso não é o mesmo que redundante. Para parâmetros irrelevantes, todas as covariâncias ou correlações, estimativas paramétrica, erros padrão, níveis de significância e graus de liberdade são configurados para o valor omissivo do sistema.

É possível usar esse arquivo de matriz como os valores iniciais para estimação de modelo adicional; observe que esse arquivo não é imediatamente utilizável para análises adicionais em outros procedimentos que leem um arquivo de matriz, a menos que esses procedimentos aceitem todos os tipos de linha exportados aqui. Mesmo assim, você deve tomar cuidado para que todos os parâmetros nesse arquivo de matriz tenham o mesmo significado para o procedimento que estiver lendo o arquivo.

**Exportar modelo como XML.** Salva as estimativas paramétrica e a matriz de covariância paramétrica, se selecionado, 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.

## Variáveis Adicionais de Comando do GENLIN

O idioma da sintaxe de comando também permite:

- Especificar valores iniciais para estimativas paramétrica como uma lista de números (usando o subcomando CRITERIA).
- Especificar uma matriz de correlação de trabalho fixa (usando o subcomando REPEATED).
- Fixar covariáveis em valores diferentes de suas médias ao calcular médias marginais estimadas (usando o subcomando EMMEANS).
- Especificar contrastes polinomiais customizados para médias marginais estimadas (usando o subcomando EMMEANS).
- Especificar um subconjunto dos fatores para os quais médias marginais estimadas são exibidas para serem comparadas usando o tipo de contraste especificado (usando as palavras-chave TABLES e COMPARE do subcomando EMMEANS).

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

## Modelos lineares generalizados mistos

Modelos lineares generalizados mistos estendem o modelo linear de forma que:



- A resposta seja linearmente relacionada com os fatores e as covariáveis através de uma função de ligação especificada.
- A resposta possa ter uma distribuição não normal.
- As observações possam ser correlacionadas.

Modelos lineares generalizados mistos cobrem uma ampla variedade de modelos, desde a regressão linear simples até os modelos complexos de vários níveis para dados longitudinais não normais.

### Exemplos

O comitê do distrito escolar pode usar um modelo linear generalizado misto para determinar se um método de ensino experimental é efetivo em melhorar as notas de matemática. Alunos da mesma sala de aula devem ser correlacionados uma vez que eles são ensinados pelo mesmo professor e salas de aula dentro da mesma escola também podem ser correlacionadas, para que possamos incluir efeitos aleatórios nos níveis escolares e de classe para ter em conta as diferentes fontes de variabilidade.

Os pesquisadores médicos podem utilizar um modelo linear generalizado misto para determinar se uma nova medicação anticonvulsiva pode reduzir a taxa de ataques epiléticos de um paciente. As medições repetidas do mesmo paciente são geralmente correlacionadas positivamente de modo que um modelo combinado com alguns efeitos aleatórios deve ser apropriado. O campo de destino, o número de convulsões, assume valores de número inteiro positivo, de modo que um modelo linear generalizado misto com distribuição de Poisson e ligação de log pode ser apropriado.

Executivos em um provedor de cabo de televisão, telefone e serviços de internet podem usar um modelo linear generalizado misto para saber mais sobre os clientes potenciais. Como possíveis respostas têm níveis de medição nominal, o analista da empresa usa um modelo logit generalizado misto com um intercepto aleatório para capturar correlação entre respostas para as perguntas de uso de serviço nos tipos de serviços (tv, telefone, internet) dentro de respostas de um respondedor de pesquisa de opinião especificado.

A guia Estrutura de Dados permite que você especifique os relacionamentos estruturais entre registros em seu conjunto de dados quando observações são correlacionadas. Se os registros no conjunto de dados representam observações independentes, não é necessário especificar nada nessa guia.

## Opções de efeitos

### Sujeitos

A combinação de valores dos campos categóricos especificados deve definir exclusivamente assuntos dentro do conjunto de dados. Por exemplo, um campo de *ID do Paciente* único deve ser suficiente para definir assuntos em um único hospital, mas a combinação de *ID do Hospital* e *ID do Paciente* pode ser necessária se números de identificação do paciente não forem exclusivos nos hospitais. Em uma configuração de medidas repetidas, várias observações são registradas para cada assunto, de modo que cada assunto pode ocupar vários registros no conjunto de dados.

Um **assunto** é uma unidade observacional que pode ser considerada independente de outros assuntos. Por exemplo, as leituras de pressão sanguínea de um paciente em um estudo médico podem ser consideradas independentes das leituras de outros pacientes. Definir assuntos torna-se particularmente importante quando existem medidas repetidas por assunto e você deseja modelar a correlação entre estas observações. Por exemplo, você pode esperar que as leituras de pressão sanguínea de um único paciente durante as visitas consecutivas para o médico sejam correlacionadas.

Todos os campos especificados como **Assuntos** no diálogo Variáveis são usados para definir assuntos para a estrutura de covariâncias residual e fornecem a lista de possíveis campos para definir assuntos para estruturas de covariâncias de efeitos aleatórios no Bloco de efeito aleatório.

### Medidas repetidas

Os campos especificados aqui são usados para identificar observações repetidas. Por exemplo, uma única variável *Semana* pode identificar as 10 semanas de observações em um estudo médico ou *Mês* e *Dia* podem ser usados juntos para identificar observações diárias no curso de um ano.

## Opções de covariância

### Definir grupos de covariância por

Os campos categóricos especificados aqui definem conjuntos independentes de parâmetros de covariância de efeitos repetidos; um para cada categoria definida pela classificação cruzada dos campos de agrupamento. Todos os assuntos têm o mesmo tipo de covariância; assuntos dentro do mesmo agrupamento de covariância terão os mesmos valores para os parâmetros.

### Tipo de covariância repetida

Isso especifica a estruturas de covariâncias para os resíduos. Diferentes opções de covariância estão disponíveis com base no **Tipo de covariância Repeada** selecionado. As estruturas disponíveis são:

- Autorregressivo de primeira ordem (AR1)
- Produto Direto AR1(UN\_AR1)
- Produto Direto não estruturado (UN\_UN)
- Simetria composta de produto direto (UN\_CS)
- Simetria composta heterogênea (CSH)
- Autorregressivo heterogêneo (ARH1)
- Média móvel autorregressiva (1,1) (ARMA11)
- Simetria composta
- Diagonal
- Identidade dimensionada
- Toeplitz
- Não estruturado
- Componente de variância
- Espacial: Energia
- Espacial: Exponencial
- Espacial: Gaussiano
- Espacial: Linear
- Espacial: Log Linear
- Espacial: Esférico

### Medidas Kronecker

Selecione variáveis que especificam a estrutura de assunto para medidas de covariância de Kronecker e determinam como os erros de medida são correlacionados. O campo estará disponível somente quando um dos seguintes **Tipos de covariância repetida** for selecionado:

- Produto Direto AR1(UN\_AR1)
- Produto Direto não estruturado (UN\_UN)
- Simetria composta de produto direto (UN\_CS)

### Coordenadas de covariância espacial

As variáveis nesta lista especificam as coordenadas das observações repetidas quando um dos tipos de covariância espacial é selecionado para o tipo de covariância repetida.

Veja o tópico [“Estruturas de Covariâncias” na página 102](#) para obter mais informações.

## Pseudo R<sup>2</sup> medidas

As medidas Pseudo-R<sup>2</sup> e o coeficiente de correlação intra-classe são incluídos na saída GLMM (quando apropriado). As medidas de pseudo-R<sup>2</sup> são baseadas inteiramente nas estimativas finais e são produzidas após a conclusão da estimativa. O coeficiente de determinação R<sup>2</sup> é uma estatística comumente relatada, pois representa a proporção de variação explicada por um modelo linear. O coeficiente de correlação

intra-classe (ICC) é uma estatística relacionada que quantifica a proporção de variância explicada por um fator de agrupamento (aleatório) em dados multiníveis / hierárquicos.

## Obtendo um modelo linear generalizado misto

Este recurso requer SPSS Statistics Edição padrão ou a opção Estatísticas avançadas.

Nos menus, escolha:

### **Analisar > Modelos Mistos > Lineares Generalizados...**

1. Defina a estrutura do assunto de seu conjunto de dados na guia **Estrutura de Dados**. A guia fornece opções para especificar os efeitos de modelo e o tipo de covariância repetida.
2. Na guia **Campos e Efeitos** deve haver uma única resposta, que pode ter qualquer nível de medição ou uma especificação de eventos/avaliações e, nesse caso, as especificações de eventos e avaliações devem ser contínuas. Opcionalmente, especifique a sua distribuição e função de ligação, os efeitos fixos e qualquer bloco de efeito aleatório, offset ou ponderação de análise.
3. Clique em **Opções de Criação** para especificar configurações opcionais de construção.
4. Clique em **Opções de Modelo** para salvar escores no conjunto de dados ativo e exportar o modelo para um arquivo externo.
5. Clique em **Executar** para executar o procedimento e criar os objetos Modelo.

## Resposta

Essas configurações definem a resposta, a sua distribuição e o seu relacionamento com os preditores por meio da função de ligação.

**Destino.** A resposta é necessária. Ela pode ter qualquer nível de medição e o nível de medição da resposta restringe quais distribuições e funções de ligação são apropriadas.

- **Usar número de avaliações como denominador.** Quando a resposta de destino for um número de eventos que ocorrem em um conjunto de avaliações, o campo de destino contém o número de eventos e é possível selecionar um campo adicional contendo o número de avaliações. Por exemplo, ao testar um novo pesticida você pode expor amostras de formigas a diferentes concentrações do pesticida e, em seguida, registrar o número de formigas mortas e o número de formigas em cada amostra. Nesse caso, o campo que registra o número de formigas mortas deve ser especificado como o campo de destino (de eventos) e o campo que registra o número de formigas em cada amostra deve ser especificado como o campo de avaliações. Se o número de formigas for o mesmo para cada amostra, então, o número de avaliações pode ser especificado usando um valor fixo.

O número de avaliações deve ser maior que ou igual ao número de eventos para cada registro. Eventos devem ser números inteiros não negativos e avaliações devem ser números inteiros positivos.

- **Customizar categoria de referência.** Para uma variável de resposta categórica, é possível escolher a categoria de referência. Isso pode afetar determinado resultado, como estimativas paramétrica, mas não deve mudar o ajuste do modelo. Por exemplo, se a sua resposta usa valores 0, 1 e 2, por padrão, o procedimento torna a última categoria (com valor mais alto) ou 2 a categoria de referência. Nessa situação, estimativas paramétrica devem ser interpretadas como relacionadas à probabilidade de categoria 0 ou 1 *relativa* à probabilidade de categoria 2. Se você especificar uma categoria customizada e a sua resposta tiver rótulos definidos, é possível configurar a categoria de referência escolhendo um valor a partir da lista. Isso pode ser conveniente quando, no meio da especificação de um modelo, você não se lembra exatamente como um determinado campo foi codificado.

**Distribuição de Resposta e Relacionamento (Link) com o Modelo Linear.** Dados os valores dos preditores, o modelo espera a distribuição de valores da resposta para seguir a forma especificada e para os valores de resposta estarem linearmente relacionados aos preditores através da função de ligação especificada. Cortes curtos para vários modelos comuns são fornecidos ou escolha uma configuração **Customizada** se houver uma determinada distribuição e a combinação de função de ligação que você deseja ajustar não estiver na lista pequena.

- **Modelo linear.** Especifica uma distribuição normal com uma ligação de identidade, que é útil quando a resposta pode ser predita usando uma regressão linear ou um modelo ANOVA.
- **Regressão Gama.** Especifica uma distribuição Gama com uma ligação de log, que deve ser usada quando a resposta contiver todos os valores positivos e for desviada em direção a valores maiores.
- **Log-Linear.** Especifica uma distribuição de Poisson com uma ligação de log, que deve ser usada quando a resposta representar uma contagem de ocorrências em um período de tempo fixo.
- **Regressão binomial negativa.** Especifica uma distribuição binomial negativa com uma ligação de log, que deve ser usada quando a resposta e o denominador representarem um número de avaliações requeridas para observar sucessos  $k$ .
- **Regressão logística multinomial.** Especifica uma distribuição multinomial, que deve ser usada quando a resposta for uma resposta de várias categorias. Ela usa uma ligação logit acumulativa (resultados ordinais) ou uma ligação logit generalizada (respostas nominais de várias categorias).
- **Regressão logística binária.** Especifica uma distribuição binomial com uma ligação logit, que deve ser usada quando a resposta for binária predita por um modelo de regressão logística.
- **Probit binário.** Especifica uma distribuição binomial com uma ligação de probito, que deve ser usada quando a resposta for binária com uma distribuição normal subjacente.
- **Sobrevivência censurada de intervalo.** Especifica uma distribuição binomial com uma ligação log-log complementar, que é útil em análise de sobrevivência quando algumas observações não têm nenhum evento de finalização.

#### Distribuição

Essa seleção especifica a distribuição da resposta. A capacidade para especificar uma distribuição não normal e a função de ligação de não identidade é a melhoria essencial do modelo linear generalizado misto sobre o modelo linear misto. Há muitas combinações de função de ligação de distribuição e diversas podem ser apropriadas para qualquer conjunto de dados especificado, de forma que a sua opção possa ser guiada por considerações teóricas a priori ou pela combinação que se ajuste melhor.

- **Binomial.** Essa distribuição é apropriada somente para uma resposta que represente uma resposta binária ou um número de eventos.
- **Gama.** Essa distribuição é apropriada para uma resposta com valores de escala positiva que são desviados em direção a valores positivos maiores. Se um valor de dados for menor que ou igual a 0 ou estiver omissa, então, o caso correspondente não será usado na análise.
- **Gaussiano Inversa.** Essa distribuição é apropriada para uma resposta com valores de escala positiva que são desviados em direção a valores positivos maiores. Se um valor de dados for menor que ou igual a 0 ou estiver omissa, então, o caso correspondente não será usado na análise.
- **Multinomial.** Essa distribuição é apropriada para uma resposta que represente uma resposta de várias categorias. A forma do modelo dependerá do nível de medição da resposta.

Uma resposta **nominal** resultará em um modelo multinomial nominal no qual um conjunto separado paramétrico de modelo é estimado para cada categoria da resposta (exceto a categoria de referência). As estimativas paramétricas para um determinado preditor mostram o relacionamento entre esse preditor e a probabilidade de cada categoria da resposta, relativo à categoria de referência.

Uma resposta **ordinal** resultará em um modelo multinomial ordinal no qual o termo de interceptação tradicional é substituído por um conjunto paramétrico de **limite** que são relacionados à probabilidade acumulativa das categorias de resposta.

- **Binomial negativa.** A regressão binomial negativa usa uma distribuição binomial negativa com uma ligação de log, que deve ser usada quando a resposta representar uma contagem de ocorrências com alta variância.
- **Normal.** Isso é apropriado para uma variável de resposta contínua cujos valores usam uma distribuição simétrica, em forma de sino sobre um valor (média) central.
- **Poisson.** Essa distribuição pode ser considerada como o número de ocorrências de um evento de interesse em um período de tempo fixo e é apropriada para variáveis com valores de número inteiro não negativo. Se um valor dos dados for não inteiro, menor que 0 ou omissa, então, o caso correspondente não será usado na análise.

## Funções de Ligação

A função de ligação é uma transformação da resposta que permite estimação do modelo. As funções a seguir estão disponíveis:

- **Identidade.**  $f(x)=x$ . A resposta não é transformada. Essa ligação pode ser usada com qualquer distribuição, exceto a multinomial.
- **Log-log complementar.**  $f(x)=\log(-\log(1-x))$ . Isso é apropriado apenas com a distribuição binomial ou multinomial.
- **Cauchit.**  $f(x) = \tan(\pi (x - 0.5))$ . Isso é apropriado apenas com a distribuição binomial ou multinomial.
- **Log.**  $f(x)=\log(x)$ . Essa ligação pode ser usada com qualquer distribuição, exceto a multinomial.
- **Complemento de log.**  $f(x)=\log(1-x)$ . Isso é apropriado apenas com a distribuição binomial.
- **Logit.**  $f(x)=\log(x / (1-x))$ . Isso é apropriado apenas com a distribuição binomial ou multinomial.
- **Log-log negativo.**  $f(x)=-\log(-\log(x))$ . Isso é apropriado apenas com a distribuição binomial ou multinomial.
- **Probit.**  $f(x)=\Phi^{-1}(x)$ , em que  $\Phi^{-1}$  é a função de distribuição acumulativa normal padrão inversa. Isso é apropriado apenas com a distribuição binomial ou multinomial.
- **Poder.**  $f(x) = x^\alpha$ , se  $\alpha \neq 0$ .  $f(x)=\log(x)$ , se  $\alpha=0$ .  $\alpha$  é a especificação numérica necessária e deve ser um número real. Essa ligação pode ser usada com qualquer distribuição, exceto a multinomial.

## Efeitos Fixos




Fatores de efeitos fixos são geralmente considerados como campos cujos valores de interesse são todos representados no conjunto de dados e podem ser usados para escoragem. Por padrão, campos com papel de entrada predefinido que não são especificados em outra parte no diálogo são inseridos na parte de efeitos fixos do modelo. Campos (flag do nominal e ordinal) categóricos são usados como fatores no modelo e campos contínuos são usados como covariáveis.

Insira efeitos no modelo, selecionando um ou mais campos na lista de origem e arrastando para a lista de efeitos. O tipo de efeito criado depende de em que ponto de acesso você solta a seleção.

- **Principal.** Campos soltos aparecem como efeitos principais separados na parte inferior da lista de efeitos.
- **bidirecional.** Todos os pares possíveis dos campos soltos aparecem como interações bidirecionais na parte inferior da lista de efeitos.
- **de três direções.** Todos os trios possíveis dos campos soltos aparecem como interações de três direções na parte inferior da lista de efeitos.
- **\***. A combinação de todos os campos soltos aparece como uma interação única na parte inferior da lista de efeitos.

Botões à direita do Construtor de Efeito permitem que você execute várias ações.

Tabela 1. Descrições do botão Construtor de Efeito

Ícone	Descrição
	Exclua termos do modelo de efeitos fixos selecionando os termos que você deseja excluir e clicando no botão Excluir.
	Reordene os termos dentro do modelo de efeitos fixos, selecionando os termos que você deseja reordenar e clicando na seta para cima ou para baixo.
	Inclua termos aninhados no modelo usando o diálogo “Incluir um Termo Customizado” na página 56, clicando no botão Incluir um Termo Customizado.

**Incluir Intercepto.** O intercepto é geralmente incluído no modelo. Se for possível assumir a passagem de dados pela origem, é possível excluir o intercepto.

## Incluir um Termo Customizado

É possível construir termos aninhados para seu modelo neste procedimento. Os termos aninhados são úteis para modelar o efeito de um fator ou covariável cujos valores não interagem com os níveis de outro fator. Por exemplo, uma rede de supermercados pode seguir os hábitos de gastos de seus clientes em vários locais de loja. Como cada cliente frequenta somente um desses locais, o efeito do *Cliente* pode ser considerado **aninhado dentro** do efeito do *Local de armazenamento*.

Além disso, é possível incluir os efeitos de interação, como termos polinomiais envolvendo a mesma covariável, ou incluir diversos níveis de aninhamento no termo aninhado.

**Limitações.** Os termos aninhados possuem as restrições a seguir:

- Todos os fatores em uma interação devem ser exclusivos. Assim, se *A* for um fator, então especificar *A\*A* será inválido.
- Todos os fatores dentro de um efeito aninhado devem ser exclusivos. Assim, se *A* for um fator, então especificar *A(A)* será inválido.
- Nenhum efeito pode ser aninhado dentro de uma covariável. Dessa forma, se *A* for um fator e *X* for uma covariável, então, especificar *A(X)* é inválido.

Construindo um termo aninhado

1. Selecione um fator ou uma covariável que seja aninhada dentro de um outro fator e, em seguida, clique no botão de seta.
2. Clique em **(Dentro)**.
3. Selecione o fator dentro do qual o fator ou a covariável anterior esteja aninhada e, em seguida, clique no botão de seta.
4. Clique em **Incluir Termo**.

Opcionalmente, é possível incluir efeitos de interação ou incluir vários níveis de aninhamento no termo aninhado.

## Efeitos Aleatórios

Fatores de efeitos aleatórios são campos cujos valores no arquivo de dados podem ser considerados uma amostra aleatória a partir de uma população maior de valores. Eles são úteis para explicar variabilidade de excesso na resposta. Por padrão, se você tiver selecionado mais de um assunto na guia Estrutura de Dados, um bloco de Efeito Aleatório será criado para cada assunto além do assunto mais profundo. Por exemplo, se você selecionou Escola, Classe e Aluno como assuntos na guia Estrutura de Dados, os blocos de efeito aleatório a seguir serão criados automaticamente:

- Efeito Aleatório 1: o assunto é escola (sem nenhum efeito, somente intercepto)
- Efeito Aleatório 2: o assunto é escola \* classe (sem nenhum efeito, somente intercepto)

É possível trabalhar com blocos de efeito aleatório nas maneiras a seguir:

1. Para incluir um novo bloco, clique em **Incluir bloco...** Isso abre o diálogo “Bloco de Efeito Aleatório” na página 56.
2. Para editar um bloco existente, selecione o bloco que deseja editar e clique em **Editar bloco...** Isso abre o diálogo “Bloco de Efeito Aleatório” na página 56.
3. Para excluir um ou mais blocos, selecione os blocos que você deseja excluir e clique no botão Excluir.

## Bloco de Efeito Aleatório

Insira efeitos no modelo selecionando um ou mais campos na lista de origem e incluindo-os na lista **Construtor de efeito**.

O tipo de efeito criado depende do **Tipo** selecionado. Campos (flag do nominal e ordinal) categóricos são usados como fatores no modelo e campos contínuos são usados como covariáveis.

### **Interação**

A combinação de todos os campos aparece como uma única interação na parte inferior da lista de efeitos.

### **Efeitos principais**

Campos soltos aparecem como efeitos principais separados na parte inferior da lista de efeitos.

### **Todos de 2 fatores**

Todos os pares possíveis dos campos soltos aparecem como interações bidirecionais na parte inferior da lista de efeitos.

### **Todos de 3 fatores**

Todos os trios possíveis dos campos soltos aparecem como interações de três direções na parte inferior da lista de efeitos.

### **Todos de 4 fatores**

Todos os possíveis trios dos campos descartados aparecem como interações de 4 vias na parte inferior da lista de efeitos.

### **Todos de 5 fatores**

Todos os possíveis trios dos campos descartados aparecem como interações de 5 vias na parte inferior da lista de efeitos.

### **Incluir intercepto**

O intercepto não é incluído no modelo de efeitos aleatórios por padrão. Se for possível assumir a passagem de dados pela origem, é possível excluir o intercepto.

### **Exibir previsões de parâmetro para este bloco**

Especifica a exibição das estimativas de parâmetro de efeitos aleatórios.

### **Combinação de assuntos**

Isso permite especificar assuntos de efeito aleatório de combinações pré-configuradas de assuntos do diálogo Variáveis. Por exemplo, se *Escola*, *Classe* e *Estudante* forem definidos como assuntos no diálogo Variáveis, e nessa ordem, a lista suspensa Combinação de assuntos terá **Nenhum**, **Escola**, **Escola \* Classe** e **Escola \* Classe \* Estudante** como opções.

### **Tipo de covariância de efeito aleatório**

Isso especifica a estruturas de covariâncias para os resíduos. As estruturas disponíveis são:

- Autorregressivo de primeira ordem (AR1)
- Média móvel autorregressiva (1,1) (ARMA11)
- Simetria composta
- Diagonal
- Identidade dimensionada
- Toeplitz
- Não estruturado
- Componente de variância

### **Definir grupos de covariância por**

Os campos categóricos especificados aqui definem conjuntos independentes de parâmetros de covariância de efeitos aleatórios; um para cada categoria definida pela classificação cruzada dos campos de agrupamento. Um conjunto diferente de campos de agrupamento pode ser especificado para cada bloco de efeito aleatório. Todos os assuntos têm o mesmo tipo de covariância; assuntos dentro do mesmo agrupamento de covariância terão os mesmos valores para os parâmetros.

## **Ponderação e Offset**

**Ponderação de Análise.** O parâmetro de escala é um parâmetro de modelo estimado relacionado à variância da resposta. As ponderações de análise são valores "conhecidos" que podem variar de observação para observação. Se o campo de ponderação de análise for especificado, o parâmetro de

escala, que está relacionado à variância da resposta, é dividido pelos valores de ponderação de análise para cada observação. Registros com valor de ponderação de análise que forem menores que ou iguais a 0 ou estiverem omissos não serão usados na análise.

**Offset.** O termo de offset é um preditor "estrutural". O seu coeficiente não é estimado pelo modelo, mas é assumido como tendo o valor 1; assim, os valores do offset são simplesmente incluídos no preditor linear da resposta. Isso é especialmente útil em modelos de regressão de Poisson, em que cada caso pode ter diferentes níveis de exposição ao evento de interesse.

Por exemplo, ao modelar as taxas de acidentes para motoristas individuais, há uma diferença importante entre um motorista que falhou em um acidente em três anos de experiência e um motorista que falhou em um acidente em 25 anos! O número de acidentes pode ser modelado como uma resposta de Poisson ou binomial negativa com uma ligação de log se o logaritmo natural da experiência do motorista estiver incluído como um termo de offset.

Outras combinações de distribuição e tipos de link iriam requerer outras transformações da variável de offset.

## Opções de Criação Geral

Essas seleções especificam mais alguns critérios avançados usados para construir o modelo.

### Ordem de classificação

Esses controles determinam a ordem das categorias para a resposta e os fatores (entradas categóricas) para propósitos de determinar a "última" categoria. A configuração de ordenação da resposta é ignorada se a resposta não for categórica ou se uma categoria de referência customizada for especificada nas configurações do ["Resposta"](#) na página 53.

### Regras de parada

É possível especificar o número máximo de iterações que o algoritmo executará. O algoritmo usa um processo duplamente iterativo que consiste em um loop interno e um loop externo. O valor que é especificado para o número máximo de iterações aplica-se a ambos os loops. Especifique um número inteiro não negativo. O padrão é 100.

### Configurações pós-estimação

Essas configurações determinam como alguma saída do modelo é calculada para visualização.

#### Nível de confiança (%)

Esse é o nível de confiança usado para calcular estimativas de intervalo dos coeficientes do modelo. Especifique um valor maior que 0 e menor que 100. O padrão é 95.

#### Graus de liberdade

Isso especifica como os graus de liberdade são calculados para testes de significância.

Escolha **Método residual** se seu tamanho da amostra for suficientemente grande, ou os dados forem balanceados, ou se o modelo usar um tipo de covariância mais simples (por exemplo, identidade escalada ou diagonal). Esta é a definição padrão. Escolha **Aproximação**

**Satterthwaite** se seu tamanho da amostra for pequeno, ou os dados não forem balanceados ou se o modelo usar um tipo de covariância complicado (por exemplo, não estruturado).

Escolha **Aproximação Kenward-Roger** se seu tamanho da amostra for pequeno e você tiver um modelo Restricted Maximum Likelihood (REML).

#### Testes de efeitos e coeficientes fixos

Esse é o método para calcular a matriz de covariâncias de estimativas de parâmetros. Escolha a estimativa robusta se você estiver preocupado que as premissas do modelo sejam violadas.

## Estimação

O algoritmo de construção de modelo usa um processo duplamente iterativo que consiste em um loop interno e um loop externo. As configurações a seguir aplicam-se ao loop interno.



**Convergência de Parâmetro.**

A convergência será assumida se a mudança absoluta máxima ou a mudança relativa máxima nas estimativas de parâmetro for menor do que o valor especificado, que deve ser não negativo. O critério não será utilizado se o valor especificado for igual a 0.

**Convergência de Log da verossimilhança**

A convergência será assumida se a mudança absoluta ou mudança relativa na função de log da verossimilhança for menor do que o valor especificado, que deve ser não negativo. O critério não será utilizado se o valor especificado for igual a 0.

**Convergência da Hessiana.**

Para uma especificação **Absoluta**, a convergência será assumida se uma estatística baseada na Hessiana for menor que o valor especificado. Para a especificação **Relativa**, a convergência será assumida se a estatística for menor que o produto do valor especificado e do valor absoluto do log da verossimilhança. O critério não será utilizado se o valor especificado for igual a 0.

**Passos de escoragem de Fisher Máxima.**

Especifique um número inteiro não negativo. Um valor 0 especifica o método de Newton-Raphson. Valores maiores que 0 especificam para usar o algoritmo de escoragem de Fisher até o número de iteração  $n$ , em que  $n$  é o número inteiro especificado e Newton-Raphson depois disso.

**Tolerância de singularidade.**

Esse valor é usado como a tolerância na verificação de singularidade. Especifique um valor positivo.

**Nota:** Por padrão, Convergência de Parâmetro é usada, em que a mudança **Absoluta** máxima em uma tolerância de 1E-6 é verificada. Essa configuração pode produzir resultados que diferem dos resultados que são obtidos em versões anteriores à versão 22. Para reproduzir resultados a partir de versões anteriores à 22, use **Relativo** para o critério de Convergência de Parâmetro e mantenha o valor de tolerância padrão 1E-6.

## Médias Estimadas

Essa guia permite que você exiba as médias marginais estimadas para níveis de fatores e interações entre fatores. Médias marginais estimadas não estão disponíveis para modelos multinomiais.

**Termos.** Os termos modelo nos Efeitos Fixos que são inteiramente compostos de campos categóricos estão listados aqui. Verifique cada termo para o qual você deseja que o modelo produza médias marginais estimadas.

- **Tipo de Contraste.** Isso especifica o tipo de contraste para usar para os níveis do campo de contraste. Se **Nenhum** for selecionado, nenhum contraste será produzido. **Entre pares** produz comparações entre pares para todas as comparações de nível dos fatores especificados. Esse é o único contraste disponível para interações entre fatores. Contrastes de **desvio** comparam cada nível do fator com a média global. Contrastes **simples** comparam cada nível do fator, exceto o último, com o último nível. O "último" nível é determinado pela ordenação para fatores especificados nas Opções de Criação. Observe que todos esses tipos de contraste não são ortogonais.
- **Campo de Contraste.** Isso especifica um fator, cujos níveis são comparados usando o tipo de contraste selecionado. Se **Nenhum** for selecionado como o tipo de contraste, nenhum campo de contraste pode (ou precisa) ser selecionado.

**Campos Contínuos.** Os campos contínuos listados são extraídos dos termos nos Efeitos Fixos que usam campos contínuos. Ao calcular médias marginais estimadas, covariáveis são fixadas nos valores especificados. Selecione a média ou especifique um valor customizado.

**Exibir médias estimadas em termos de.** Isso especifica se deve-se calcular médias marginais estimadas com base na escala original da resposta ou com base na transformação da função de ligação. A **Escala de resposta original** calcula médias marginais estimadas para a resposta. Observe que a resposta é especificada usando a opção de eventos/avaliações; isso fornece as médias marginais estimadas para a proporção de eventos/avaliações ao invés de para o número de eventos. A **Transformação da função de ligação** calcula médias marginais estimadas para o preditor linear.

**Ajustar para várias comparações usando.** Ao executar testes de hipótese com vários contrastes, o nível de significância global pode ser ajustado a partir dos níveis de significância para os contrastes incluídos. Isso permite que você escolha o método de ajuste.

- **Diferença menos significativa.** Esse método não controla a probabilidade geral de rejeitar a hipótese de que algum contraste linear seja diferente dos valores de hipótese nula.
- *Bonferroni Sequencial.* Este é um procedimento de Bonferroni de rejeição sequencialmente decrescente que tende ser muito menos conservador em termos de rejeição de hipóteses individuais, mas mantém o mesmo nível de significância geral.
- *Sidak Sequencial.* Este é um procedimento de Sidak de rejeição sequencialmente decrescente que tende ser muito menos conservador em termos de rejeição de hipóteses individuais, mas mantém o mesmo nível de significância geral.

O método de diferença menos significativa é menos conservador que o método Sidak sequencial, que por sua vez é menos conservador do que o Bonferroni sequencial; ou seja, a diferença menos significativa rejeitará pelo menos tantas hipóteses individuais quantas Sidak sequenciais, que por sua vez rejeitarão pelo menos tantas hipóteses individuais quantas Bonferroni sequenciais.

## Salvar

Itens marcados são salvos com o nome especificado; conflitos com nomes de campo existentes não são permitidos.

**Valores preditos-** salva o valor predito da resposta. O nome do campo padrão é *PredictedValue*.

**Probabilidade predita para destinos categóricos-** Se a resposta for categórica, essa palavra-chave salvará as probabilidades preditivas das primeiras *n* categorias, até o valor especificado como **Máximo de categorias a serem salvas**. Os valores calculados são probabilidades acumulativas para respostas ordinais. O nome raiz padrão é *PredictedProbability*. Para salvar a probabilidade predita da categoria predita, salve a confiança (consulte abaixo).

**Intervalos de confiança-** Salva os limites superior e inferior do intervalo de confiança para o valor predito ou probabilidade predita. Para todas as distribuições exceto a multinomial, isso cria duas variáveis e o nome raiz padrão é *CI*, com *\_Lower* e *\_Upper* como os sufixos.

Para a distribuição multinomial e uma resposta nominal, um campo é criado para cada categoria de variável dependente. Isso salva os limites superiores e inferiores da probabilidade predita para as primeiras categorias *n* até o valor especificado como as **Categorias máximas a serem salvas**. O nome raiz padrão é *CI* e os nomes de campo padrão são *CI\_Lower\_1*, *CI\_Upper\_1*, *CI\_Lower\_2*, *CI\_Upper\_2*, e assim por diante, correspondendo à ordem das categorias de destino.

Para a distribuição multinomial e uma resposta ordinal, um campo é criado para cada categoria de variável dependente exceto a última (Consulte o tópico “Opções de Criação Geral” na página 58 para obter mais informações.). Isso salva os limites superiores e inferiores da probabilidade predita acumulativa para as primeiras categorias *n*, até, mas não incluindo a última, o valor especificado como as **Categorias máximas a serem salvas**. O nome raiz padrão é *CI* e os nomes de campo padrão são *CI\_Lower\_1*, *CI\_Upper\_1*, *CI\_Lower\_2*, *CI\_Upper\_2*, e assim por diante, correspondendo à ordem das categorias de destino.

**Resíduos de Pearson-** salva o resíduo de Pearson para cada registro, que pode ser usado em diagnósticos pós-estimação do ajuste do modelo. O nome do campo padrão é *PearsonResidual*.

**Confianças-** Salva a confiança no valor predito para a resposta categórica. A confiança calculada pode ser baseada na probabilidade do valor predito (a probabilidade predita mais alta) ou na diferença entre a probabilidade predita mais alta e a segunda a probabilidade predita mais alta. O nome do campo padrão é *Confiança*.

## Exportar

**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. Consulte o tópico Assistente de Pontuação para obter mais informações. Especifique um nome do arquivo

exclusivo, válido. Se a especificação do arquivo se referir a um arquivo existente, então, o arquivo será sobrescrito.

### Exportar saída

**Exportar EBLUPS.**Essa seção será ativada se pelo menos um bloco de efeitos aleatórios for criado por meio do diálogo do item Efeitos Aleatórios na guia Campos e Efeitos e a caixa de seleção para exibir previsões de parâmetros para esse bloco for selecionada. Se vários desses blocos tiverem sido especificados, será possível optar por ter os resultados salvos em conjuntos de dados ou arquivos separados ou combinados em um único conjunto de dados ou arquivo.

## Visualização do modelo

O procedimento cria um objeto Modelo no Visualizador. Ao ativar (clique duas vezes) nesse objeto, você obtém uma visualização interativa do modelo.

Por padrão, a visualização de Sumarização do Modelo é mostrada. Para ver uma outra visualização do modelo, selecione-a a partir das miniaturas de visualização.

Como alternativa ao objeto Modelo, é possível gerar tabelas e gráficos dinâmicos ao selecionar **Tabelas e gráficos dinâmicos** no grupo Exibição de saída na guia Saída do diálogo Opções (Editar > Opções). Os tópicos que seguem descrevem o objeto Modelo.

### Sumarização do Modelo

Essa visualização é uma captura instantânea, sumarização de visão rápida do modelo e de seu ajuste.

**Tabela.** A tabela identifica a resposta, distribuição de probabilidade e função de ligação especificadas nas Configurações de destino. Se a resposta for definida por eventos e avaliações, a célula será dividida para mostrar o campo de eventos e o campo de avaliações ou o número fixo de avaliações. Além disso, o critério de informações de Akaike corrigido de amostra finita (AICC) e o critério de informações Bayesiano (BIC) são exibidos.

- *Akaike corrigido.* Uma medida para selecionar e comparar modelos mistos com base na probabilidade de logaritmo -2 (restrita). Valores menores indicam melhores modelos. O AICC "corrige" o AIC para tamanhos de amostra pequenos. Conforme o tamanho da amostra aumenta, o AICC converge para o AIC.
- *Bayesiano.* Uma medida para selecionar e comparar modelos com base na probabilidade de registro -2. Valores menores indicam melhores modelos. O BIC também "penaliza" os modelos sobreparametrizados (modelos complexos com um grande número de entradas, por exemplo), mas é mais rígido do que o AIC.

**Gráfico.** Se a resposta for categórica, um gráfico exibe a precisão do modelo final, que é a porcentagem de classificações corretas.

### Estrutura de Dados

Essa visualização fornece uma sumarização da estrutura de dados que você especificou e ajuda a verificar se os assuntos e as medidas repetidas foram especificados corretamente. As informações observadas para o primeiro assunto são exibidas para cada campo de assunto e campo de medidas repetidas, e a resposta. Além disso, o número de níveis para cada campo de assunto e campo de medidas repetidas é exibido.

### Predito por Observado

Para variáveis de resposta contínua, incluindo respostas especificadas como eventos/avaliações, isso exibe um gráfico de dispersão categorizado dos valores preditos 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 previsto de forma inválida pelo modelo.

## Classificação

Para variáveis resposta categórica, isso exibe a classificação cruzada de valores observados versus preditos em um heat map, mais o percentual geral correto.

**Estilos da tabela.** Existem vários estilos de exibição diferentes, que são acessíveis a partir da lista suspensa de **Estilos**.

- **Percentuais de linhas.** Isso exibe as porcentagens de linhas (as contagens de célula expressas como um percentual dos totais de linhas) nas células. Esse é o padrão.
- **Contagens de célula.** Isso exibe as contagens de célula nas células. O sombreamento para o heat map ainda é baseado nas porcentagens de linhas.
- **Heat map.** Isso não exibe nenhum valor nas células, apenas o sombreamento.
- **Compactado.** Isso não exibe nenhum título da linha ou da coluna ou valor nas células. Isso pode ser útil quando a resposta tem muitas categorias.

**Omissão.** Se qualquer registro tiver valores omissos na resposta, ele será exibido em uma linha (**Omissão**) sob todas as linhas válidas. Registros com valores omissos não contribuem com o percentual geral correto.

**Várias respostas.** Se houver múltiplas variáveis resposta categórica, então, cada resposta será exibida em uma tabela separada e há uma lista suspensa de **Respostas** que controla qual resposta deve ser exibida.

**Tabelas grandes.** Se a resposta exibida tiver mais de 100 categorias, nenhuma tabela será exibida.

## Efeitos Fixos

Essa visualização exibe o tamanho de cada efeito fixo 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 na ordem na qual eles foram especificados nas configurações de Efeitos Fixos. Linhas de conexão no diagrama são ponderadas com base na significância do efeito, com maior largura da linha correspondendo a efeitos mais significativos (valores  $p$  menores). Esse é o padrão.
- **Tabela.** Essa é uma tabela ANOVA para o modelo global e os efeitos do modelo individuais. Os efeitos individuais são ordenados da parte superior até a inferior na ordem na qual eles foram especificados nas configurações de Efeitos Fixos.

**Significância.** Há uma Régua de controle de significância que controla quais efeitos são mostrados na visualização. Efeitos com valores de significância maiores que o valor da régua de controle são ocultados. Isso não muda o modelo, mas simplesmente permite que você se concentre nos efeitos mais importantes. Por padrão o valor é 1,00, de forma que nenhum efeito seja filtrado com base na significância.

## Coeficientes Fixos

Essa visualização exibe o valor de cada coeficiente fixo 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 coeficiente 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 na ordem na qual eles foram especificados nas configurações de Efeitos Fixos. Dentro de efeitos que contêm fatores, coeficientes são ordenados por ordem ascendente de valores dos dados. Linhas de conexão no diagrama são coloridas e ponderadas com base na significância do coeficiente, com maior largura da linha correspondendo a coeficientes mais significativos (valores  $p$  menores). Esse é o estilo padrão.

- **Tabela.** Isso mostra os valores, testes de significância e intervalos de confiança para os coeficientes do modelo individuais. Após o intercepto, os efeitos são ordenados da parte superior até a inferior na ordem na qual eles foram especificados nas configurações de Efeitos Fixos. Dentro de efeitos que contêm fatores, coeficientes são ordenados por ordem ascendente de valores dos dados.

**Multinomial.** Se a distribuição multinomial estiver em vigor, então, a lista suspensa Multinomial controla qual categoria de destino deve ser exibida. A ordenação dos valores na lista é determinada pela especificação nas configurações de Opções de Criação.

**Exponencial.** Isso exibe estimativas de coeficiente exponenciais e intervalos de confiança para determinados tipos de modelo, incluindo Regressão logística binária (distribuição binomial e link logit), Regressão logística nominal (distribuição multinomial e link logit), Regressão logística negativa (distribuição binomial negativa e ligação de log), e Modelo log-linear (distribuição de Poisson e ligação de log).

**Significância.** Há uma Régua de controle de significância que controla quais coeficientes são mostrados na visualização. Coeficientes com valores de significância maiores que o valor da régua de controle são ocultados. Isso não muda o modelo, mas simplesmente permite que você se concentre nos coeficientes mais importantes. Por padrão o valor é 1,00, de forma que nenhum coeficiente seja filtrado com base na significância.

## Covariâncias de Efeito Aleatório

Essa visualização exibe a matriz de covariâncias de efeitos aleatórios (**G**).

**Estilos.** Existem estilos de exibição diferentes, que são acessíveis a partir da lista suspensa de **Estilos**.

- **Valores de covariância.** Esse é um heat map da matriz de covariâncias no qual os efeitos são ordenados da parte superior até a inferior na ordem na qual eles foram especificados nas configurações de Efeitos Fixos. As cores no corrgram correspondem aos valores da célula conforme mostrado na chave. Esse é o padrão.
- **Corrgram.** Esse é um heat map da matriz de covariâncias.
- **Compactado.** Esse é um heat map da matriz de covariâncias sem os títulos da linha e da coluna.

**Blocos.** Se houver vários blocos de efeito aleatório, então, há uma lista suspensa de Blocos para selecionar o bloco a ser exibido.

**Grupos.** Se um bloco de efeito aleatório tiver uma especificação de grupo, então, há uma lista suspensa de Grupos para selecionar o nível do grupo a ser exibido.

**Multinomial.** Se a distribuição multinomial estiver em vigor, então, a lista suspensa Multinomial controla qual categoria de destino deve ser exibida. A ordenação dos valores na lista é determinada pela especificação nas configurações de Opções de Criação.

## Parâmetros de Covariância

Essa visualização exibe as estimativas paramétrica de covariância e estatísticas relacionadas para efeitos de resíduo e aleatórios. Esses são resultados avançados, mas fundamentais, que fornecem informações sobre se a estruturas de covariâncias é adequada.

**Tabela de sumarização.** Essa é uma referência rápida para o número paramétricas nas matrizes de covariâncias de efeito de resíduo (**R**) e aleatório (**G**), o ranqueamento (número de colunas) nas matrizes de design de objeto fixo (**X**) e de efeito aleatório (**Z**) e o número de assuntos definidos pelos campos de assunto que definem a estrutura de dados.

**Tabela paramétrica de covariância.** Para o efeito selecionado, a estimativa, o erro padrão e o intervalo de confiança são exibidos para cada parâmetro de covariância. O número paramétricas mostrados depende da estruturas de covariâncias para o efeito e, para blocos de efeito aleatório, o número de efeitos no bloco. Se você vir que os parâmetros fora da diagonal não são significativos, você pode ser capaz de usar uma estruturas de covariâncias mais simples.

**Efeitos.** Se houver blocos de efeito aleatório, então, há uma lista suspensa de Efeitos para selecionar o bloco de efeito residual ou aleatório a ser exibido. O efeito de resíduo está sempre disponível.

**Grupos.** Se um bloco de efeito de resíduo ou aleatório tiver uma especificação de grupo, então, há uma lista suspensa de Grupos para selecionar o nível do grupo a ser exibido.

**Multinomial.** Se a distribuição multinomial estiver em vigor, então, a lista suspensa Multinomial controla qual categoria de destino deve ser exibida. A ordenação dos valores na lista é determinada pela especificação nas configurações de Opções de Criação.

## Médias Estimadas: Efeitos Significativos

Esses são gráficos exibidos para os 10 efeitos de todos os fatores fixos "mais significativos", iniciando com as interações de três fatores e, em seguida, as interações de dois fatores e, finalmente, efeitos principais. O gráfico exibe o valor estimado pelo modelo da resposta no eixo vertical para cada valor do efeito principal (ou primeiro efeito listado em uma interação) no eixo horizontal; uma linha separada é produzida para cada valor do segundo efeito listado em uma interação; um gráfico separado é produzido para cada valor do terceiro efeito listado em uma interação de três fatores; todos os outros preditores são mantidos constantes. Ele fornece uma visualização útil dos efeitos dos coeficientes de cada preditor na resposta. Observe que se nenhum preditor for significativo, nenhuma média estimada será produzida.

**Confiança.** Isso exibe limites de confiança superiores e inferiores para as médias marginais, usando o nível de confiança especificado como parte das Opções de Criação.

## Médias Estimadas: Efeitos Customizados

Essas são tabelas e gráficos para efeitos de todos os fatores solicitados pelo usuário.

**Estilos.** Existem estilos de exibição diferentes, que são acessíveis a partir da lista suspensa de **Estilos**.

- **Diagrama.** Este estilo exibe um gráfico de linha do valor estimado pelo modelo da resposta no eixo vertical para cada valor do efeito principal (ou primeiro efeito listado em uma interação) no eixo horizontal; uma linha separada é produzida para cada valor do segundo efeito listado em uma interação; um gráfico separado é produzido para cada valor do terceiro efeito listado em uma interação de três fatores; todos os outros preditores são mantidos constantes.

Se contrastes foram solicitados, outro gráfico é exibido para comparar níveis do campo de contraste; para interações, um gráfico é exibido para cada combinação de nível dos efeitos diferente do campo de contraste. Para contrastes **entre pares**, ele é um gráfico de rede de distância; ou seja, uma representação gráfica da tabela de comparações na qual as distâncias entre nós na rede correspondem a diferenças entre amostras. Linhas amarelas correspondem a diferenças estatisticamente significativas; linhas pretas correspondem a diferenças não significativas. Passar o mouse sobre uma linha na rede exibe uma dica de ferramenta com a significância ajustada da diferença entre os nós conectados pela linha.

Para contrastes de **desvio**, um gráfico de barras é exibido com o valor estimado pelo modelo da resposta no eixo vertical e os valores do campo de contraste no eixo horizontal; para interações, um gráfico é exibido para cada combinação de nível dos efeitos diferente do campo de contraste. As barras mostram a diferença entre cada nível do campo de contraste e a média geral, que é representada por uma linha horizontal preta.

Para contrastes **simples**, um gráfico de barras é exibido com o valor estimado pelo modelo da resposta no eixo vertical e os valores do campo de contraste no eixo horizontal; para interações, um gráfico é exibido para cada combinação de nível dos efeitos diferente do campo de contraste. As barras mostram a diferença entre cada nível do campo de contraste (exceto o último) e o último nível, que é representada por uma linha horizontal preta.

- **Tabela.** Esse estilo exibe uma tabela do valor estimado pelo modelo da resposta, o seu erro padrão e o intervalo de confiança para cada combinação de nível dos campos no efeito; todos os outros preditores são mantidos constantes.

Se contrastes foram solicitados, outra tabela é exibida com a estimativa, o erro padrão, o teste de significância e o intervalo de confiança para cada contraste; para interações, há um conjunto separado de linhas para cada combinação de nível dos efeitos diferente do campo de contraste. Além disso, uma tabela com os resultados gerais de teste é exibida; para interações, há um teste geral separado para cada combinação de nível dos efeitos diferente do campo de contraste.

**Confiança.** Isso alterna a exibição de limites de confiança superiores e inferiores para as médias marginais, usando o nível de confiança especificado como parte das Opções de Criação.

**Layout.** Isso alterna o layout do diagrama de contrastes entre pares. O layout de círculo é menos revelador de contrastes do que o layout de rede, mas evita linhas de sobreposição.

## Análise de log linear de seleção do modelo

---

O procedimento de Análise Log-Linear de Seleção de Modelo analisa tabulações cruzadas de vários caminhos (tabelas de contingência). Ele ajusta modelos log-linear hierárquicos a tabulações cruzadas multidimensionais usando um algoritmo de ajuste proporcional iterativo. Esse procedimento o ajuda a descobrir quais variáveis categóricas estão associadas. Para construir modelos, métodos de entrada forçada e de eliminação backward estão disponíveis. Para modelos saturados, é possível solicitar estimativas paramétrica e testes de associação parcial. Um modelo saturado inclui 0,5 em todas as células.

**Exemplo.** Em um estudo de preferência do usuário para um dos dois detergentes para roupas, pesquisadores contaram pessoas em cada grupo, combinando várias categorias de suavidade da água (suave, média ou difícil), uso prévio de uma das marcas e temperatura de lavagem (fria ou quente). Eles descobriram como a temperatura é relacionada à suavidade da água e também à preferência de marca.

**Estatísticas.** Frequências, resíduos, estimativas paramétrica, erros padrão, intervalos de confiança e testes de associação parcial. Para modelos customizados, gráficos de resíduos e gráficos de probabilidade normal.

Considerações de Dados de Análise Log-Linear de Seleção de Modelo

**Dados.** Variáveis de fator são categóricas. Todas as variáveis a serem analisadas devem ser numéricas. Variáveis de sequência de caracteres categóricas podem ser registradas para variáveis numéricas antes de iniciar a análise de seleção de modelo.

Evite especificar muitas variáveis com níveis. Essas especificações podem levar a uma situação em que muitas células têm pequenos números de observações e os valores de qui-quadrado podem não ser úteis.

**Procedimentos relacionados.** O procedimento de Seleção de Modelo pode ajudar a identificar os termos necessários no modelo. Em seguida, é possível continuar a avaliar o modelo usando Análise Log-Linear Geral ou Análise Log-Linear Logit. É possível usar Recodificação Automática para recodificar variáveis de sequência de caracteres. Se uma variável numérica tiver categorias vazias, use Recodificar para criar valores de número inteiro consecutivos.

Obtendo uma Análise Log-Linear de Seleção de Modelo

Nos menus, escolha:

**Analisar > Log-Linear > Seleção de Modelo...**

1. Selecione dois ou mais fatores categóricos numéricos.
2. Selecione uma ou mais variáveis de fator na lista de Factores e clique em **Definir Intervalo**.
3. Defina o intervalo de valores para cada variável de fator.
4. Selecione uma opção no grupo de Construção de Modelo.

Opcionalmente, é possível selecionar uma variável de ponderação de célula para especificar zeros estruturais.

## Análise Log-Linear - Definir Intervalo

Deve-se indicar o intervalo de categorias para cada variável de fator. Valores para Mínimo e Máximo correspondem às categorias mais baixa e mais alta da variável de fator. Ambos os valores devem ser números inteiros e o valor mínimo deve ser inferior ao valor máximo. Casos com valores fora dos limites são excluídos. Por exemplo, se você especificar um valor mínimo de 1 e um valor máximo de 3, apenas os valores 1, 2 e 3 serão usados. Repita este processo para cada variável de fator.

## Modelo de Análise Log-Linear

**Especificar modelo.** Um modelo saturado contém todos os efeitos principais do fator e todas as interações de fator por fator. Selecione **customizado** para especificar uma classe geradora para um modelo não saturado.

**Classe Geradora.** Uma classe geradora é uma lista de termos de ordem mais alta nos quais fatores aparecem. Um modelo hierárquico contém os termos que definem a classe geradora e todos os parentes de ordem inferior. Suponha que você selecione variáveis  $A$ ,  $B$  e  $C$  na lista de Fatores e, em seguida, selecione **Interação** a partir da lista suspensa Construir Termos. O modelo resultante conterá a interação de 3 fatores especificadas  $A*B*C$ , as interações de 2 fatores  $A*B$ ,  $A*C$  e  $B*C$ , e efeitos principais para  $A$ ,  $B$  e  $C$ . Não especifique os relativos de ordem mais baixa na classe geradora.

Para os fatores selecionados:

### Interação

Cria o termo de interação de nível mais alto de todas as variáveis selecionadas. Esse é o padrão.

### Efeitos principais

Cria um termo dos principais efeitos para cada variável selecionada.

### Todos de 2 fatores

Cria todas as interação de dois fatores das variáveis selecionadas.

### Todos de 3 fatores

Cria todas as interações de três fatores possíveis das variáveis selecionadas.

### Todos de 4 fatores

Cria todas as interações de quatro fatores possíveis das variáveis selecionadas.

### Todos de 5 fatores

Cria todas as interações de cinco fatores possíveis das variáveis selecionadas.

## Termos de construção e termos customizados

### Construir termos

Use esta opção quando quiser incluir termos não aninhados de um tipo específico (como efeitos principais) para todas as combinações de um conjunto selecionado de fatores e covariáveis.

### Construir termos customizados

Use esta opção quando quiser incluir termos aninhados ou quando quiser construir explicitamente qualquer variável de termo por variável. A construção de um termo aninhado envolve as seguintes etapas:

## Opções de Análise Log-Linear de Seleção de Modelo

**Exibição.** É possível escolher **Frequências**, **Resíduos** ou ambos. Em um modelo saturado, as frequências observadas e esperadas são iguais e os resíduos são iguais a 0.

**Gráfico.** Para modelos customizados, é possível escolher um ou ambos os tipos de gráficos, **Resíduos** e **Probabilidade Normal**. Esses ajudarão a determinar quão bem um modelo ajusta os dados.

**Exibição para Modelo Saturado.** Para um modelo saturado, é possível escolher **Estimativas paramétrica**. As estimativas paramétrica podem ajudar a determinar quais termos podem ser descartados do modelo. Uma tabela de associação, que lista testes de associação parcial, também está disponível. Essa opção é computacionalmente dispendiosa para tabelas com muitos fatores.

**Crítérios de Modelo.** Um algoritmo de ajuste proporcional iterativo é usado para obter estimativas paramétrica. É possível substituir um ou mais dos critérios de estimação especificando **Iterações máximas**, **Convergência** ou **Delta** (a valor incluído em todas as frequências de célula para modelos saturados).

## Recursos Adicionais do Comando HILOGLINEAR

O idioma da sintaxe de comando também permite:



- Especifique ponderações de célula no formulário matriz (usando o subcomando CWEIGHT).
- Gere análises e vários modelos com um único comando (usando o subcomando DESIGN).

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

## Análise de log linear geral

O procedimento de Análise Log-Linear Geral analisa as contagens de frequência de observações que caem em cada categoria de classificação cruzada em uma tabulação cruzada ou uma tabela de contingência. Cada classificação cruzada na tabela constitui uma célula e cada variável categórica é chamada de um fator. A variável dependente é o número de casos (frequência) em uma célula da tabulação cruzada e as variáveis explanatórias são fatores e covariáveis. Esse procedimento estima parâmetros de máxima verossimilhança de modelos log-lineares hierárquicos e não hierárquicos usando o método de Newton-Raphson. Uma distribuição de Poisson ou multinomial pode ser analisada.

É possível selecionar até 10 fatores para definir as células de uma tabela. Uma variável de estrutura de célula permite definir zeros estruturais para tabelas incompletas, incluir um termo de offset no modelo, ajustar um modelo de log-taxa ou implementar o método de ajuste de tabelas marginais. Variáveis de contraste permitem cálculo de razões de possibilidades de log generalizadas (GLOR).

Informações de modelo e estatísticas de qualidade de ajuste são automaticamente exibidas. Também é possível exibir uma variedade de estatísticas e gráficos ou salvar resíduos e valores preditos no conjunto de dados ativo.

**Exemplo.** Dados de um relatório de acidentes de automóvel na Flórida são usados para determinar o relacionamento entre o uso de um cinto de segurança e se um dano foi fatal ou não fatal. A razão de chances indica evidência significativa de um relacionamento.

**Estatísticas.** Frequências observadas e esperadas; resíduos brutos, ajustados e de deviance; matriz de design; estimativas paramétrica; razão de chances; razão de possibilidades de log; GLOR; estatística de Wald; e intervalos de confiança. Gráficos: resíduos ajustados, resíduos de deviance e probabilidade normal.

Considerações de Dados de Análise Log-Linear Geral

**Dados.** Fatores são categóricos e covariáveis de célula são contínuas. Quando uma covariável estiver em um modelo, o valor da covariável média para casos em uma célula é aplicado naquela célula. Variáveis de contraste são contínuas. Elas são usadas para calcular razões de possibilidades de log generalizadas. Os valores da variável de contraste são os coeficientes para a combinação linear dos logs das contagens de células esperadas.

Uma variável de estrutura de célula atribui ponderações. Por exemplo, se algumas das células forem zeros estruturais, a variável de estrutura de célula tem um valor 0 ou 1. Não use uma variável de estrutura de célula para ponderar dados agregados. Em vez disso, escolha **Casos de Ponderação** a partir do menu Dados.

**Suposições.** Duas distribuições estão disponíveis na Análise Log-Linear Geral: de Poisson e multinomial.

Sob a suposição de distribuição de Poisson:

- O tamanho total da amostra não é fixo antes do estudo ou a análise não é condicional no tamanho total da amostra.
- O evento de uma observação estar em uma célula é estatisticamente independente das contagens de células de outras células.

Sob a suposição de distribuição multinomial:

- O tamanho total da amostra é fixo ou a análise é condicional no tamanho total da amostra.
- As contagens de células não são estatisticamente independentes.

**Procedimentos relacionados.** Use o procedimento de Tabela Cruzada para examinar as tabulações cruzadas. Use o procedimento Log-Linear Logit quando for natural considerar uma ou mais variáveis categóricas como as variáveis de resposta e as outras como as variáveis explanatórias.

Obtendo uma Análise Log-Linear Geral

1. Nos menus, escolha:

**Analisar > Log-Linear > Geral...**

2. Na caixa de diálogo Análise Log-Linear Geral, selecione até 10 variáveis de fator.

Como opção, você pode:

- Selecione covariáveis de célula.
- Selecione uma variável de estrutura de célula para definir zeros estruturais ou incluir um termo de offset.
- Selecione uma variável de contraste.

## Modelo de Análise Log-Linear Geral

**Especificar modelo.** Um modelo saturado contém todos os efeitos e interações principais que envolvem variáveis de fator. Ele não contém termos de covariável. Selecione **Customizado** para especificar apenas um subconjunto de interações ou para especificar interações fator por covariável.

**Fatores & covariáveis.** Os fatores e covariáveis são listados.

**Termos no Modelo.** O modelo depende da natureza de seus dados. Após selecionar **Customizado**, é possível selecionar os efeitos e interações principais que forem de interesse em sua análise. Deve-se indicar todos os termos a serem incluídos no modelo.

Para os fatores e covariáveis selecionados:

### Interação

Cria o termo de interação de nível mais alto de todas as variáveis selecionadas. Esse é o padrão.

### Efeitos principais

Cria um termo dos principais efeitos para cada variável selecionada.

### Todos de 2 fatores

Cria todas as interações de dois fatores das variáveis selecionadas.

### Todos de 3 fatores

Cria todas as interações de três fatores possíveis das variáveis selecionadas.

### Todos de 4 fatores

Cria todas as interações de quatro fatores possíveis das variáveis selecionadas.

### Todos de 5 fatores

Cria todas as interações de cinco fatores possíveis das variáveis selecionadas.

## Termos de construção e termos customizados

### Construir termos

Use esta opção quando quiser incluir termos não aninhados de um tipo específico (como efeitos principais) para todas as combinações de um conjunto selecionado de fatores e covariáveis.

### Construir termos customizados

Use esta opção quando quiser incluir termos aninhados ou quando quiser construir explicitamente qualquer variável de termo por variável. A construção de um termo aninhado envolve as seguintes etapas:

## Opções de Análise Log-Linear Geral

O procedimento de Análise Log-Linear Geral exibe informações e estatísticas de qualidade de ajuste de modelo. Além disso, é possível escolher um ou mais dos seguintes:

**Exibição.** Várias estatísticas estão disponíveis para exibição - frequências de célula observadas e esperadas; resíduos brutos, ajustados e de deviance; uma matriz de design do modelo; e estimativas paramétrica para o modelo.

**Gráfico.** Gráficos, disponíveis apenas para modelos customizados, incluem duas matrizes de gráfico de dispersão (resíduos ajustados ou resíduos de deviance com relação a contagens de células observadas e esperadas). Também é possível exibir probabilidade normal e gráficos normais retificados de resíduos ajustados ou resíduos de deviance.

**Intervalo de Confiança.** O intervalo de confiança para estimativas paramétrica pode ser ajustado.

**Crítérios.** O método Newton-Raphson é utilizado para obter estimativas de parâmetros de máxima verossimilhança. É possível inserir novos valores para o número máximo de iterações, o critério de convergência e o delta (uma constante incluída em todas as células para aproximações iniciais). O delta permanece nas células para modelos saturados.

## Salvamento de Análise Log-Linear Geral

Selecione os valores que você deseja salvar como novas variáveis no conjunto de dados ativo. O sufixo *n* nos nomes da nova variável é incrementado para criar um nome exclusivo para cada variável salva.

Os valores salvos se referem aos dados agregados (células na tabela de contingência), mesmo se os dados forem registrados em observações individuais no Editor de Dados. Se você salvar valores de resíduos ou preditos para dados não agregados, o valor salvo para uma célula na tabela de contingência será inserido no Editor de Dados para cada caso naquela célula. Para fazer sentido dos valores salvos, você deve agregar os dados para obter as contagens de célula.

Quatro tipos de resíduos podem ser salvos: bruto, padronizado, ajustado e deviance. Os valores preditos ser podem ser salvos.

- *Resíduos.* Também chamado de resíduo simples ou bruto, é a diferença entre a contagem de células observadas e sua contagem esperada.
- *Resíduos padronizados.* O residual dividido por uma estimativa de seu erro padrão. Resíduos padronizados também são conhecidos como resíduos de Pearson.
- *Resíduos ajustados.* O resíduo padronizado dividido por seu erro padrão estimado. Como os resíduos ajustados são assintoticamente um padrão normal quando o modelo selecionado está correto, eles estão preferencialmente acima dos resíduos padronizados para verificar a normalidade.
- *Resíduos de deviance.* A raiz quadrada assinada de uma contribuição individual para a estatística qui-quadrado da razão de verossimilhança (G quadrado), em que o sinal é o sinal do resíduo (contagem observada menos contagem esperada). Os resíduos de deviance possuem uma distribuição normal padrão assintótica.

## Recursos Adicionais do Comando GENLOG

O idioma da sintaxe de comando também permite:

- Calcular combinações lineares de frequências de célula observadas e esperadas e imprimir resíduos, resíduos padronizados e resíduos ajustados dessa combinação (usando o subcomando GERESID).
- Alterar o valor do limite padrão para a verificação de redundância (utilizando o subcomando CRITERIA).
- Exibir os resíduos padronizados (utilizando o subcomando PRINT).

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

## Análise Log-Linear Logit

O procedimento de Análise Log-Linear Logit analisa o relacionamento entre variáveis dependentes (ou de resposta) e independentes (ou explanatórias). As variáveis dependentes são sempre categóricas, enquanto as variáveis independentes podem ser categóricas (fatores). Outras variáveis independentes (covariáveis de célula) podem ser contínuas, mas elas não são aplicadas em uma base de caso a caso. A média de covariável ponderada para uma célula é aplicada a essa célula. O logaritmo das chances das variáveis dependentes é expresso como uma combinação linear paramétricas. Uma distribuição multinomial é automaticamente considerada; esses modelos às vezes são chamados de modelos logit

multinomiais. Esse procedimento estima parâmetros de modelos log-lineares logit usando o algoritmo de Newton-Raphson.

É possível selecionar de 1 a 10 variáveis dependentes e de fator combinadas. Uma variável de estrutura de célula permite definir zeros estruturais para tabelas incompletas, incluir um termo de offset no modelo, ajustar um modelo de log-taxa ou implementar o método de ajuste de tabelas marginais.. Variáveis de contraste permitem cálculo de razões de possibilidades de log generalizadas (GLOR). Os valores da variável de contraste são os coeficientes para a combinação linear dos logs das contagens de células esperadas.

Informações de modelo e estatísticas de qualidade de ajuste são automaticamente exibidas. Também é possível exibir uma variedade de estatísticas e gráficos ou salvar resíduos e valores preditos no conjunto de dados ativo.

**Exemplo.** Um estudo na Flórida incluiu 219 jacarés. Como o tipo de alimentação de jacarés varia com o seu tamanho e os quatro lagos em que eles vivem? O estudo detectou que as chances de um jacaré menor preferir répteis para pescar é 0,70 vezes menor que para jacarés maiores; além disso, as chances de selecionar primeiramente répteis ao invés de peixe foram mais altas no lago 3.

**Estatísticas.** Frequências observadas e esperadas; resíduos brutos, ajustados e de deviance; matriz de design; estimativas paramétrica; razão de possibilidades de log generalizadas; estatística de Wald; e intervalos de confiança. Gráficos: resíduos ajustados, resíduos de deviance e gráficos de probabilidade normal.

#### Considerações de Dados de Análise Log-Linear Logit

**Dados.** As variáveis dependentes são categóricas. Os fatores são categóricos. Covariáveis de célula podem ser contínuas, mas quando uma covariável está no modelo, o valor da covariável média para casos em uma célula é aplicado naquela célula. Variáveis de contraste são contínuas. Elas são usadas para calcular razões de possibilidades de log generalizadas (GLOR). Os valores da variável de contraste são os coeficientes para a combinação linear dos logs das contagens de células esperadas.

Uma variável de estrutura de célula atribui ponderações. Por exemplo, se algumas das células forem zeros estruturais, a variável de estrutura de célula tem um valor 0 ou 1. Não use uma variável de estrutura de célula para ponderar dados agregados. Em vez disso, use Casos de Ponderação no menu Dados.

**Suposições.** As contagens dentro de cada combinação de categorias de variáveis explanatórias são consideradas como tendo uma distribuição multinomial. Sob a suposição de distribuição multinomial:

- O tamanho total da amostra é fixo ou a análise é condicional no tamanho total da amostra.
- As contagens de células não são estatisticamente independentes.

**Procedimentos relacionados.** Use o procedimento de Tabela Cruzada para exibir as tabelas de contingência. Use o procedimento de Análise Log-Linear Geral quando deseja analisar o relacionamento entre uma contagem observada e um conjunto de variáveis explanatórias.

#### Obtendo uma Análise Log-Linear Logit

1. Nos menus, escolha:

**Analisar > Log-Linear > Logit...**

2. Na caixa de diálogo Análise Log-Linear Logit, selecione uma ou mais variáveis dependentes.
3. Selecione uma ou mais variáveis de fator.

O número total de variáveis dependentes e de fator deve ser inferior ou igual a 10.

Como opção, você pode:

- Selecione covariáveis de célula.
- Selecione uma variável de estrutura de célula para definir zeros estruturais ou incluir um termo de offset.
- Selecionar uma ou mais variáveis de contraste.

## Modelo de Análise Log-Linear Logit

**Especificar modelo.** Um modelo saturado contém todos os efeitos e interações principais que envolvem variáveis de fator. Ele não contém termos de covariável. Selecione **Customizado** para especificar apenas um subconjunto de interações ou para especificar interações fator por covariável.

**Fatores & covariáveis.** Os fatores e covariáveis são listados.

**Termos no Modelo.** O modelo depende da natureza de seus dados. Após selecionar **Customizado**, é possível selecionar os efeitos e interações principais que forem de interesse em sua análise. Deve-se indicar todos os termos a serem incluídos no modelo.

Para os fatores e covariáveis selecionados:

### Interação

Cria o termo de interação de nível mais alto de todas as variáveis selecionadas. Esse é o padrão.

### Efeitos principais

Cria um termo dos principais efeitos para cada variável selecionada.

### Todos de 2 fatores

Cria todas as interações de dois fatores das variáveis selecionadas.

### Todos de 3 fatores

Cria todas as interações de três fatores possíveis das variáveis selecionadas.

### Todos de 4 fatores

Cria todas as interações de quatro fatores possíveis das variáveis selecionadas.

### Todos de 5 fatores

Cria todas as interações de cinco fatores possíveis das variáveis selecionadas.

Termos são incluídos no design tomando todas as combinações possíveis dos termos dependentes e comparando cada combinação com cada termo na lista de modelos. Se **Incluir constante para dependente** for selecionado, também há um termo de unidade (1) incluído na lista de modelos.

Por exemplo, suponha que as variáveis  $D1$  e  $D2$  sejam as variáveis dependentes. Uma lista de termos dependentes é criada pelo procedimento de Análise Log-Linear Logit ( $D1$ ,  $D2$ ,  $D1*D2$ ). Se os Termos na lista de Modelos contiverem  $M1$  e  $M2$  e uma constante for incluída, a lista de modelos conterá 1,  $M1$  e  $M2$ . O design resultante inclui combinações de cada termo modelo com cada termo dependente:

$D1$ ,  $D2$ ,  $D1*D2$

$M1*D1$ ,  $M1*D2$ ,  $M1*D1*D2$

$M2*D1$ ,  $M2*D2$ ,  $M2*D1*D2$

**Incluir constante para dependente.** Inclui uma constante para a variável dependente em um modelo customizado.

## Termos de construção e termos customizados

### Construir termos

Use esta opção quando quiser incluir termos não aninhados de um tipo específico (como efeitos principais) para todas as combinações de um conjunto selecionado de fatores e covariáveis.

### Construir termos customizados

Use esta opção quando quiser incluir termos aninhados ou quando quiser construir explicitamente qualquer variável de termo por variável. A construção de um termo aninhado envolve as seguintes etapas:

## Opções de Análise Log-Linear Logit

O procedimento de Análise Log-Linear Logit exibe informações e estatísticas de qualidade de ajuste de modelo. Além disso, é possível escolher uma ou mais das opções a seguir:

**Exibição.** Várias estatísticas estão disponíveis para exibição: frequências de célula observadas e esperadas; resíduos brutos, ajustados e de deviance; uma matriz de design do modelo; e estimativas paramétrica para o modelo.

**Gráfico.** Gráficos disponíveis para modelos customizados incluem duas matrizes de gráfico de dispersão (resíduos ajustados ou resíduos de deviance com relação a contagens de células observadas e esperadas). Também é possível exibir probabilidade normal e gráficos normais retificados de resíduos ajustados ou resíduos de deviance.

**Intervalo de Confiança.** O intervalo de confiança para estimativas paramétrica pode ser ajustado.

**Crítérios.** O método Newton-Raphson é utilizado para obter estimativas de parâmetros de máxima verossimilhança. É possível inserir novos valores para o número máximo de iterações, o critério de convergência e o delta (uma constante incluída em todas as células para aproximações iniciais). O delta permanece nas células para modelos saturados.

## Salvamento de Análise Log-Linear Logit

Selecione os valores que você deseja salvar como novas variáveis no conjunto de dados ativo. O sufixo *n* nos nomes da nova variável é incrementado para criar um nome exclusivo para cada variável salva.

Os valores salvos se referem aos dados agregados (a células na tabela de contingência), mesmo se os dados forem registrados em observações individuais no Editor de Dados. Se você salvar valores de resíduos ou preditos para dados não agregados, o valor salvo para uma célula na tabela de contingência será inserido no Editor de Dados para cada caso naquela célula. Para fazer sentido dos valores salvos, você deve agregar os dados para obter as contagens de célula.

Quatro tipos de resíduos podem ser salvos: bruto, padronizado, ajustado e deviance. Os valores preditos ser podem ser salvos.

- *Resíduos.* Também chamado de resíduo simples ou bruto, é a diferença entre a contagem de células observadas e sua contagem esperada.
- *Resíduos padronizados.* O residual dividido por uma estimativa de seu erro padrão. Resíduos padronizados também são conhecidos como resíduos de Pearson.
- *Resíduos ajustados.* O resíduo padronizado dividido por seu erro padrão estimado. Como os resíduos ajustados são assintoticamente um padrão normal quando o modelo selecionado está correto, eles estão preferencialmente acima dos resíduos padronizados para verificar a normalidade.
- *Resíduos de deviance.* A raiz quadrada assinada de uma contribuição individual para a estatística qui-quadrado da razão de verossimilhança (G quadrado), em que o sinal é o sinal do resíduo (contagem observada menos contagem esperada). Os resíduos de deviance possuem uma distribuição normal padrão assintótica.

## Recursos Adicionais do Comando GENLOG

O idioma da sintaxe de comando também permite:

- Calcular combinações lineares de frequências de célula observadas e esperadas e imprimir resíduos, resíduos padronizados e resíduos ajustados dessa combinação (utilizando o subcomando GERESID).
- Alterar o valor do limite padrão para a verificação de redundância (utilizando o subcomando CRITERIA).
- Exibir os resíduos padronizados (utilizando o subcomando PRINT).

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

## Tabelas de mortalidade

Há muitas situações nas quais você gostaria de examinar a distribuição de vezes entre dois eventos, como duração de emprego (tempo entre ser contratado e deixar a empresa). No entanto, esse tipo de dados usualmente inclui alguns casos para os quais o segundo evento não está registrado (por exemplo, pessoas que ainda trabalham para a empresa no final do estudo). Isto pode acontecer por várias razões:

para alguns casos, o evento simplesmente não ocorre antes do final do estudo; para outros casos, perdemos o controle de seu status algum tempo antes do final do estudo; ainda em outros casos pode não ser possível continuar por razões não relacionadas ao estudo (como um empregado adoecendo e tirando uma licença). Coletivamente, esses casos são conhecidos como **casos censurados** e eles tornam esse tipo de estudo inapropriado para técnicas tradicionais como testes  $t$  ou regressão linear.

Um técnica estatística útil para esse tipo de dados é chamada de **tabela de mortalidade** de acompanhamento. A ideia básica da tabela de mortalidade é subdividir o período de observação em intervalos de tempo menores. Para cada intervalo, todas as pessoas que foram observadas pelo menos por aquele tempo são usadas para calcular a probabilidade de um evento terminal ocorrer nesse intervalo. As probabilidades estimadas a partir de cada um dos intervalos são, então, usadas para estimar a probabilidade geral do evento ocorrer em pontos de tempo diferentes.

**Exemplo.** Uma nova terapia de adesivo de nicotina é melhor do que a terapia de adesivo tradicional para ajudar as pessoas a pararem de fumar? Você poderia realizar um estudo usando dois grupos de fumantes, um do quais receberia a terapia tradicional e o outro receberia a terapia experimental. Construir tabelas de mortalidade a partir dos dados lhe permitiria comparar taxas de abstinência globais entre os dois grupos para determinar se o tratamento experimental é uma melhoria em relação à terapia tradicional. Também é possível criar gráfico de funções de sobrevivência ou de risco e compará-las visualmente para obter informações mais detalhadas.

**Estatísticas.** Inserção de número, número ignorado, número exposto a risco, número de eventos terminais, finalização de proporção, sobrevivência à proporção, sobrevivência à proporção acumulativa (e erro padrão), densidade de probabilidade (e erro padrão) e taxa de risco (e erro padrão) para cada intervalo de tempo para cada grupo; tempo de sobrevivência de mediana para cada grupo; e teste de Wilcoxon (Gehan) para comparar distribuições de sobrevivência entre grupos. Gráficos: gráficos de função para sobrevivência, sobrevivência de log, densidade, taxa de risco e um menos a sobrevivência.

Considerações de Dados de Tabelas de Mortalidade

**Dados.** A sua variável de tempo deve ser quantitativa. A sua variável de status deve ser dicotômica ou categórica, codificada como números inteiros, com eventos sendo codificados como um valor único ou um intervalo de valores consecutivos. Variáveis de fator devem ser categóricas, codificadas como números inteiros.

**Suposições.** Probabilidades para o evento de interesse devem depender apenas do tempo após o evento inicial--elas são consideradas como estáveis com relação ao tempo absoluto. Isso é, casos que entram no estudo em tempos diferentes (por exemplo, pacientes que começaram o tratamento em tempos diferentes) devem se comportar de forma semelhante. Também não deve haver diferenças sistemáticas entre casos censurados e não censurados. Se, por exemplo, muitos dos casos censurados forem pacientes com condições mais sérias, os seus resultados podem ser influenciados.

**Procedimentos relacionados.** O procedimento de Tabelas de Mortalidade usa uma abordagem atuarial para esse tipo de análise (conhecida geralmente como Análise de Sobrevivência). O procedimento de Análise de Sobrevivência de Kaplan-Meier usa um método ligeiramente diferente para calcular tabelas de mortalidade que não dependem do particionamento do período de observação em intervalos de tempo menores. Esse método é recomendado se você tiver um número pequeno de observações, de modo que haveria apenas um número pequeno de observações em cada intervalo de tempo de sobrevivência. Se você tiver variáveis que suspeita estarem relacionadas a tempo de sobrevivência ou variáveis que você deseja controlar (covariáveis), use o procedimento de Regressão de Cox. Se as suas covariáveis podem ter valores diferentes em pontos diferentes no tempo para o mesmo caso, use Regressão de Cox com Covariáveis Dependentes de Tempo.

Criando Tabelas de Mortalidade

1. Nos menus, escolha:

**Analisar > Sobrevivência > Tabelas de Mortalidade...**

2. Selecione uma variável de sobrevivência *numérica*.

3. Especifique os intervalos de tempo a serem examinados.

4. Selecione uma variável de status para definir casos para os quais o evento terminal ocorreu.

5. Clique em **Definir Evento** para especificar o valor da variável de status que indica que um evento ocorreu.

Opcionalmente, é possível selecionar uma variável de fator de primeira ordem. Tabelas atuariais para a variável de sobrevivência são geradas para cada categoria da variável de fator.

Também é possível selecionar uma variável de *by fator* de segunda ordem. Tabelas atuariais para a variável de sobrevivência são geradas para cada combinação das variáveis de fator de primeira e segunda ordem.

## Definir Intervalo para Tabelas de Mortalidade

Casos com valores para a variável de fator no intervalo que você especificar serão incluídos na análise, e tabelas separadas (e gráficos, se solicitado) serão geradas para cada valor exclusivo no intervalo.

## Opções de Tabelas de Mortalidade

É possível controlar vários aspectos da análise de Tabelas de Mortalidade.

**Tabela(s) de mortalidade.** Para suprimir a exibição de tabelas de mortalidade na saída, cancele a seleção de **Tabela(s) de mortalidade**.

**Gráfico.** Permite solicitar gráficos das funções de sobrevivência. Se você tiver uma ou mais variáveis de fator, os gráficos serão gerados para cada subgrupo definido por uma ou mais variáveis de fator. Os gráficos disponíveis são de sobrevivência, log de sobrevivência, risco, densidade e um menos a sobrevivência.

- *Sobrevivência.* Exibe a função de sobrevivência acumulativa em uma escala linear.
- *Sobrevivência de log.* Exibe a função de sobrevivência acumulativa em uma escala logarítmica.
- *Risco.* Exibe a função de risco acumulativo em uma escala linear.
- *Densidade.* Exibe a função de densidade.
- *Um menos sobrevivência.* Exibe a função um menos a sobrevivência em uma escala linear.

**Comparar Níveis do Primeiro Fator.** Se você tiver uma variável de controle de primeira ordem, será possível selecionar uma das alternativas neste grupo para executar o teste de Wilcoxon (Gehan), que compara a sobrevivência de subgrupos. Os testes são executados no fator de primeira ordem. Se tiver definido um fator de segunda ordem, os testes serão executados para cada nível da variável de segunda ordem.

## Variáveis Adicionais de Comando do SURVIVAL

O idioma da sintaxe de comando também permite:

- Especificar mais de uma variável dependente.
- Especificar intervalos espaçados desigualmente.
- Especificar mais de uma variável de status.
- Especificar comparações que não incluam todas as variáveis de fator e de controle.
- Calcular comparações aproximadas, ao invés de exatas.

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

## Modelos De Tempo De Falha Acelerada Paramétrica

Uma Análise de Modelo de Falha Acelerada Paramétrica (AFT) invoca o procedimento de modelos de sobrevivência paramétrica com dados de tempo de vida não recorrentes. Os modelos de sobrevivência paramétrica supõem que o tempo de sobrevivência siga uma distribuição conhecida, e esta análise se encaixa em modelos de tempo de falha acelerada com seus efeitos de modelo proporcionais com relação ao tempo de sobrevivência.



## Obtenção de uma análise de Modelos de Tempo de Falha Acelerada Paramétrica

1. Nos menus, escolha:

**Analisar > Survival > Modelos de Falha Acelerada Paramétrica (AFT) Models**

2. Selecione uma variável de origem.

### Hora

#### Sobrevivência

Única variável numérica que denota a duração do tempo de sobrevivência.

#### Início/Término

Variáveis numéricas denotando **Hora de Início** e **Horário de Fim**.

### Estado

Única string opcional ou variável numérica que determina uma das configurações de status a seguir:

#### Falha/Evento

Mapeia um registro para uma categoria de falha / evento. O valor padrão para uma variável de status string é F.

#### Censura à direita

Mapeia um registro para uma categoria de censura à direita. O valor padrão para uma variável de status string é R.

#### Censura à esquerda

Mapeia um registro para uma categoria de censura à esquerda. O valor padrão para uma variável de status string é L.

#### Censura de intervalo

Mapeia um registro para uma categoria de censura de intervalo. Para **Iniciar / Fim** apenas. O valor padrão para uma variável de status string é I.

#### Tratamento de valores não mapeados

Controla a qual categoria mapear os registros não mapeados para. Para excluir os registros que não conseguiram ser mapeados, selecione **Excluí-los da análise**.

Para **Survival**, o status padrão para todos os casos é **Falha / Evento**. Para **Início / Fim**, o status padrão é **Censoring Interval**. Clique no botão **Definir evento** para definir um evento para a variável de status.

### Covariável(is)

Uma ou mais variáveis numéricas opcionais a serem tratadas como covariados. Note que uma variável não pode ser especificada por ambos **Covariate (s)** e **Fixed Factor (s)**.

### Fatores fixos

Uma ou mais variáveis opcionais a serem tratadas como fatores. Uma variável não pode ser especificada por ambos **Fator Fixo (s)** e **Covariado (s)**.

### Truncamento à esquerda

Única variável numérica opcional para truncamento esquerdo para **Survival** apenas.

## Modelos De Tempo De Falha Acelerada Paramétrica: Critérios

### Critérios

Um painel opcional para especificar os critérios gerais.

#### Intervalo de confiança

Uma porcentagem opcional para especificar o nível para os intervalos de confiança dos parâmetros de regressão. Ele deve ser um valor numérico único entre 0 e 100. O padrão é 95.

#### Valores omissos

Uma opção para controlar como os valores ausentes do usuário são tratados:

#### Excluir valores omissos do usuário e do sistema

Trata os valores ausentes do usuário como valores válidos. Esse é o padrão.

**Valores omissos para o usuário são tratados como válidos**

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

**Tratamento de status**

Para **Iniciar / Fim** apenas. Uma opção para controlar como lidar com registros com campos de status incorretos:

**Descartar registro conflituado**

Cai os registros conflituados. Esta é a definição padrão.

**Obter as informações de horário de acordo com o status**

Obtém as informações de tempo de acordo com o status.

**Derivar o status de acordo com as informações de tempo**

Altera o status de acordo com as informações do tempo.

## Modelos De Tempo De Falha Acelerada Paramétrica: Modelo

**Modelo**

Um painel opcional para especificar as opções e configurações do modelo.

**Distribuição do tempo de sobrevivência**

Uma opção para especificar a distribuição do tempo de sobrevivência.

**Weibull**

Especifica a distribuição Weibull. Esta é a definição padrão.

**Exponencial**

Especifica distribuição exponencial.

**Log-normal**

Especifica log-distribuição normal.

**Logística de log**

Especifica distribuição log-logistic.

**Configurações de covariável**

Especificar variáveis covariadas.

**Configurações do fator**

Especificar variáveis de fator.

**Valor inicial de interceptação**

Uma opção para especificar o valor inicial do termo de interceptação. Se especificado, ele deve ser um único valor numérico, e não pode ser 0.

**Valor inicial do parâmetro de escala**

Uma opção para controlar a configuração do parâmetro de escala.

**Erro padrão da regressão OLS correspondente**

Usa o erro padrão da regressão ordinária de mínimos quadrados correspondente como o valor inicial.

**Inverte erro padrão da regressão OLS correspondente**

Usa o recíproco do erro padrão.

**Valor fornecido pelo usuário**

Se um único valor numérico for especificado, o valor será usado como o valor inicial. Se especificado, deve ser maior que 0.

## Modelos De Tempo De Falha Acelerada Paramétrica: Estimativa

**Estimativa**

Um painel opcional para especificar as configurações para controlar a estimação dos modelos de tempo de falha acelerada e o processo de seleção de recurso opcional.

## Método De Direção Alternada ou Multiplicadores (ADMM)

### Rápido

Aplica o método de direção de alternância rápida de multiplicadores (ADMM). Esse é o padrão.

### Tradicional

Aplica o algoritmo ADMM tradicional.

## Aplicar regularização L-1

Conduz o processo para controlar a seleção de recursos. O Campo **Penalty Parâmetro** especifica o parâmetro de pênalti que controla o processo de regularização. Deve ser um valor único maior que 0. A configuração padrão é 0.001..

## CrITÉRIOS de convergência modelo

### Convergência de parâmetro

Especifica os critérios de convergência para o parâmetro. Deve ser um valor numérico único pertencente a  $[0, 1]$ . A configuração padrão é 0.000001. Para **Tipo**, você pode selecionar **ABSOLUTO** para aplicar a convergência absoluta à otimização interna ou **RELATIVA** para aplicar a convergência relativa à otimização interna. O campo **Valor** opcional especifica uma palavra-chave.

### Convergência de função objetiva

Especifica os critérios de convergência para a função objetiva. Deve ser um valor numérico único pertencente a  $[0, 1]$ . A configuração padrão é de 0, o que não aplica os critérios de convergência. Para **Tipo**, você pode selecionar **ABSOLUTO** para aplicar a convergência absoluta à otimização interna ou **RELATIVA** para aplicar a convergência relativa à otimização interna. O campo **Valor** opcional especifica uma palavra-chave.

### Convergência Hessiana

Especifica os critérios de convergência para a matriz Hessiana. Deve ser um valor numérico único pertencente a  $[0, 1]$ . A configuração padrão é de 0, o que não aplica os critérios de convergência. Para **Tipo**, você pode selecionar **ABSOLUTO** para aplicar a convergência absoluta à otimização interna ou **RELATIVA** para aplicar a convergência relativa à otimização interna. O campo **Valor** opcional especifica uma palavra-chave.

## CrITÉRIOS de convergência residual

Uma opção para controlar o processo de otimização.

### Resíduos primários e duplos

Aplica critério de convergência residual primal e dual. Esta é a definição padrão.

### Apenas residual primário

Aplica o critério de convergência residual primal.

### Apenas residual duplo

Aplica o critério de convergência residual dual.

## Método

Um parâmetro opcional para especificar o método de estimação.

### Automático

Escolhe automaticamente o método baseado no conjunto de dados de amostra. Esse é o padrão. O Campo **Limiar de número de preditores** especifica o limite do número de preditores e deve ser um único inteiro maior que 1. O valor padrão é 1000.

### Newton-Raphson

Aplica o método de Newton-Raphson.

### L-BFGS

Aplica o algoritmo BFGS de memória limitada. O campo **Update** especifica o número das atualizações passadas mantidas pelo algoritmo BFGS de memória limitada, e deve ser um único inteiro maior ou igual a 1. O valor padrão é 5 segundos.

## Iteração

### Máximo de iterações

Especifica o número máximo de iterações. Ele deve ser um único inteiro pertencente a  $[1, 100]$ . A configuração padrão é 20.

**Máxima divisão de passos pela metade**

Especifica o número máximo de halving de etapa. Ele deve ser um único inteiro pertencente a [1, 20]. A configuração padrão é 5.

**Número máximo de pesquisas de linha**

Especifica o número máximo de pesquisas de linha. Ele deve ser um único inteiro pertencente a [1, 100]. A configuração padrão é 20.

**Convergência absoluta para o processo de iteração**

Especifica a convergência absoluta para o processo de iteração exterior. Deve ser um valor numérico único pertencente a (0, 1). A configuração padrão é 0.0001..

**Convergência relativa para o processo de iteração**

Especifica a convergência relativa para o processo de iteração externa. Deve ser um valor numérico único pertencente a (0, 1). A configuração padrão é 0.01..

## Modelos De Tempo De Falha Acelerada Paramétrica: Imprimir

### Imprimir

Um painel opcional para controlar as saídas de mesa.

**Detalhes de codificação do fator**

Se selecionado, exibe e imprime os detalhes de codificação dos fatores. O processo é ignorado se não houver fatores de efetivação.

**Valores iniciais atribuídos aos parâmetros de regressão**

Se selecionado, exibe os valores iniciais utilizados no processo de estimação.

**Histórico de iteração do modelo**

Se selecionado, exibe o histórico de iteração de análise de sobrevivência. No campo **Número de etapas**, especifique o número de etapas entre 1 e 99999999. A definição padrão é 1.

**Resultados da seleção contendo**

Controla a exibição dos detalhes da seleção do recurso.

**Ambas as variáveis selecionadas e não selecionadas**

Exibir ambas as variáveis selecionadas e não selecionadas na tabela.

**Apenas variáveis selecionadas**

Apenas exibir as variáveis selecionadas.

**Apenas variáveis não selecionadas**

Apenas exibir as variáveis não selecionadas. O campo **Variáveis máximas a exibir** especifica o número máximo das variáveis impressas na tabela. A configuração padrão é 30.

## Modelos De Tempo De Falha Acelerada Paramétrica: Prever

### Prever

Um painel opcional para pontuar e salvar as estatísticas previstas para o conjunto de dados ativos.

**Valores de tempo para escoragem****Valores de tempo definidos por variáveis dependentes**

Pontue as **Previsões** com base na variável de tempo especificada para o modelo de sobrevivência paramétrica.

**Intervalos regulares**

Pontue as **Previsões** com base em valores de tempo futuros. O campo **Intervalo de tempo** especifica o intervalo de tempo, e deve ser um valor numérico único maior que 0. O campo **Número de períodos de tempo** especifica o número dos períodos de tempo, e deve ser um único inteiro numérico entre 2 e 100.

### **Tempo de duração**

Pontue as **Previsões** com base no tempo de duração para definir os valores de tempo futuro. Ele deve ser uma única variável numérica.

### **Predições**

#### **Sobrevivência**

Marca e salva as estatísticas de sobrevivência previstas para o conjunto de dados ativos. O nome da variável customizada padrão (ou nome-raiz) é PredSurvival.

#### **Risco**

Marca e salva os perigos previstos para o conjunto de dados ativos. O nome da variável customizada padrão (ou nome-raiz) é PredHazard.

#### **Risco acumulativo**

Pontue e salva os perigos acumulativos previstos para o conjunto de dados ativos. O nome da variável customizada padrão (ou nome-raiz) é PredCumHazard.

#### **Sobrevivência condicional**

Marca e salva as estatísticas de sobrevivência condicional previstas para o conjunto de dados ativos. O nome da variável customizada padrão (ou nome-raiz) é PredConditionalSurvival. O processo será ignorado se PASTTIME não for especificado. Um valor **Tempo de sobrevivência passado** é necessário e especifica os valores de tempo passado para escoragem. Ele deve ser uma única variável numérica.

## **Modelos De Tempo De Falha Acelerada Paramétrica: Plot**

### **Gráfico**

#### **Gráfico de resíduos de Cox-Snell**

Selecione **Exibir o gráfico** para criar um gráfico residual Cox -Snell. No campo **Número de pontos de corte de binning**, especifique um número de 1 10000. A configuração padrão é 100.

#### **Gráficos de função**

Uma opção para controlar as tramas de função.

#### **Tipo**

##### **Sobrevivência**

Cria a trama para funções de sobrevivência.

##### **Risco**

Cria a trama para as funções de risco.

##### **Densidade**

Cria uma trama para as funções de densidade.

#### **Número de pontos a serem exibidos**

Especifica o número de pontos de função entre 1 e 200. A configuração padrão é 100.

#### **Valores covariados para gráfico**

Um opcional para especificar os valores fornecidos pelo usuário e atribuí-los aos preditores. Por padrão, as parcelas designadas serão criadas no **Mean** de cada covariado em efeito e na frequência de categoria de cada fator em vigor. Se especificado, as parcelas designadas serão criadas com base na configuração do padrão. Na presença de quaisquer variáveis duplicadas, o especificado primeiro seria reconhecido e o restante seria ignorado. Uma variável válida deve ser contida em um efeito modelo. Para um covariado, o valor fornecido pelo usuário deve ser numérico. Omissão de uma variável em efeito indica que a frequência da categoria e o **Mean** seriam usados por padrão para o fator e o covariado, respectivamente. Se um valor inválido for atribuído a uma variável, o padrão solicitado não será plotado.

#### **Valores de fator para gráfico**

Um opcional para especificar os valores fornecidos pelo usuário e atribuí-los aos preditores. Na presença de quaisquer variáveis duplicadas, o especificado primeiro seria reconhecido e o restante seria ignorado. Uma variável válida deve ser contida em um efeito modelo. Omissão

de uma variável em efeito indica que a frequência da categoria e a média seriam utilizadas por inadimplência para o fator e o covariado, respectivamente. Se um valor inválido for atribuído a uma variável, o padrão solicitado não será plotado.

#### **Linhas separadas para**

Uma opção para especificar uma variável categórica pela qual as parcelas da linha serão desenhadas.

#### **Número máximo de linhas em um gráfico**

Especifica o número máximo das linhas em um gráfico se **Separar linhas para** for especificado. A configuração padrão é 10.

## **Modelos De Tempo De Falha Acelerada Paramétrica: Exportação**

### **Exportar**

Selecione **Exportar informações do modelo para arquivo XML** para gravar as informações do modelo e do parâmetro em um arquivo PMML para pontuação. Você deve especificar o diretório e o nome do arquivo do arquivo PMML a ser salvo.

## **Sobrevivência AFT Definir Eventos para Variáveis de Status**

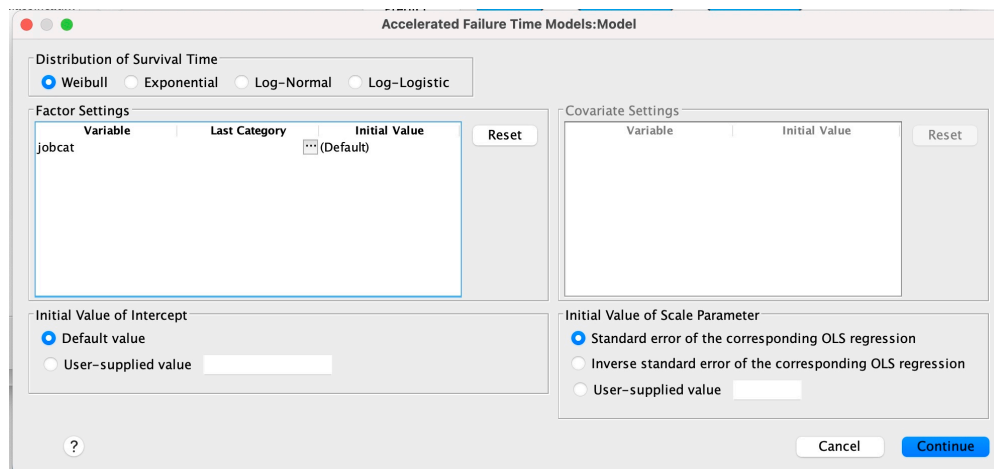
Ocorrências do valor ou valores selecionados para a variável de status indicam que o evento terminal ocorreu para esses casos. Todos os outros casos são considerados como censurados. Insira um valor único ou um intervalo de valores que identifica o evento de interesse.

## **Modelos de Tempo de Falha Acelerada Paramétrica: Selecionar Categoria**

A configuração Selecionar Categoria fornece opções para escolher um valor que indica a categoria a ser modelada como uma linha de base para comparação.

Selecione a categoria

Clique em 'Última Categoria' para abrir a caixa de diálogo 'Selecionar categoria'.



*Figura 3. Modelos de Tempo de Vida Acelerado-Caixa de Diálogo-Categoria*

Para designar uma categoria como a linha de base, selecione um valor na caixa de diálogo 'Selecionar categoria'.

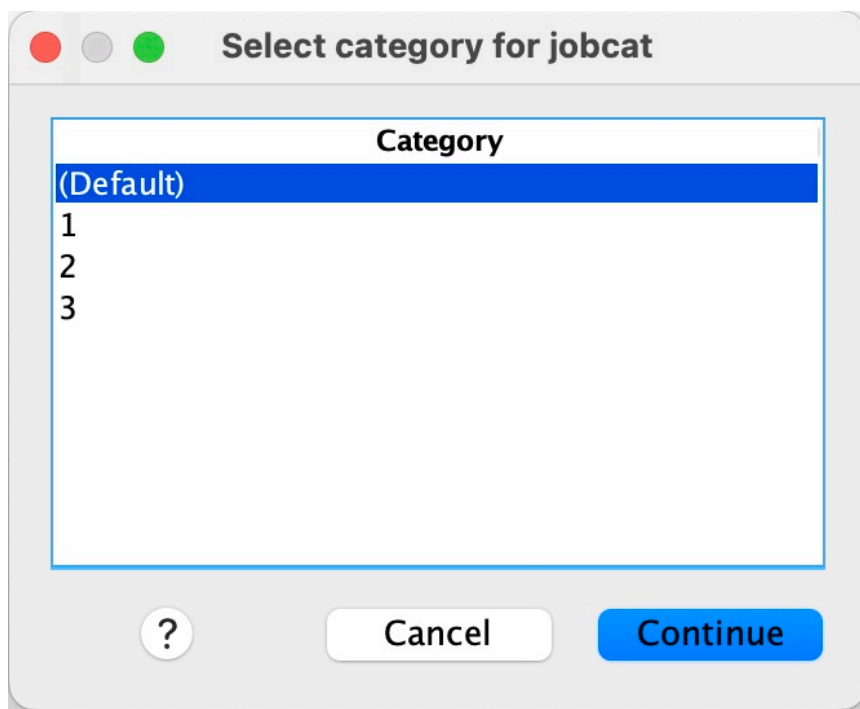


Figura 4. Modelos de Tempo de Vida Acelerado-Caixa de Diálogo-Selecionar categoria

Clique em Continuar.

## Modelos de fragilidade compartilhados paramétricos

A Parametric Shared Frailty Models Survival analysis inicia o procedimento paramétrico de modelos de sobrevivência com entrada de dados de tempo de vida recorrente. Os modelos paramétricos de sobrevivência assumem que o tempo de sobrevivência segue uma distribuição conhecida, e esta análise incorpora um termo de fragilidade em um modelo paramétrico de sobrevivência. É tratado como um componente aleatório para dar conta de um efeito não observado devido à variabilidade de nível individual ou de grupo.

### Obtendo uma análise paramétrica de Modelos de Fragilidade Compartilhados

1. No menu, escolha:

**Analisar > Sobrevivência > Modelos Paramétricos Compartilhados de Fragilidade**

2. Selecione uma variável de origem.

#### Hora

##### Sobrevivência

O tempo de sobrevivência é representado por uma variável para indicar o horário de encerramento. O horário de início seria configurado como 0

##### Início / Fim

Variáveis numéricas que denotam **Horário de Início** e **Horário de Encerramento**.

#### Assunto

Necessário para executar o procedimento.. Especifica uma única variável para o ID do sujeito..

#### Intervalo

Especifica uma variável única e numérica para o número de intervalo usado para identificar os diferentes registros recorrentes que compartilham o mesmo ID de sujeito.

**Estado**

Sequência opcional única ou variável numérica que determina uma das seguintes configurações de status:

**Falha/Evento**

Mapeia um registro para uma categoria de falha / evento O valor padrão para uma variável de status de sequência é F.

**Censura à direita**

Mapeia um registro para uma categoria de censura direita. O valor padrão para uma variável de status de sequência é R.

**Tratamento de valores não mapeados**

Controla para qual categoria mapear os registros não mapeados.. Para excluir os registros que falharam ao serem mapeados, selecione **Excluir deles da análise**.

Clique no botão **Definir evento** para definir um evento para a variável de status

**Covariável(is)**

Uma ou mais variáveis numéricas opcionais a serem tratadas como covariáveis Observe que uma variável não pode ser especificada por **Covariáveis** e **Fator (s) Fixo (s)**.

**Fatores fixos**

Uma ou mais variáveis opcionais a serem tratadas como fatores.. Uma variável não pode ser especificada por **Fator (s) Fixo (s)** e **Covariáveis**.

## Modelos de Fragilidade Compartilhados Paramétricos: Critérios

**Critérios**

Um painel opcional para especificar os critérios gerais

**Intervalo de confiança**

Uma porcentagem opcional para especificar o nível para os intervalos de confiabilidade de parâmetros de regressão Ele deve ser um único valor numérico entre 0 e 100. O padrão é 95.

**Nível de Significância**

Uma opção para especificar o nível de significância do teste de razão de verossimilhança para o componente de fragilidade Ele deve ser um valor numérico único entre 0 e 1. A configuração padrão é 0.05..

**Valores omissos**

Uma opção para controlar como os valores omissos de usuário são tratados:

**Excluir valores omissos do usuário e do sistema**

Trata os valores omissos de usuário como valores válidos.. Esse é o padrão.

**Valores omissos para o usuário são tratados como válidos**

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

**Tratamento de intervalo**

Uma opção para controlar como lidar com registros cujo intervalo entra em conflito com o horário de início e de término Ele entrará em vigor se houver duas variáveis de tempo com uma variável Intervalo especificada no diálogo principal.

**Descartar os registros com conflitos**

Descarta os registros seriais do sujeito inteiro se o valor do intervalo estiver em conflito com o horário de início e de término Esta é a definição padrão.

**Descobrir valores de intervalo com base no horário de início e término**

Descobre o valor do intervalo do horário de início e de encerramento.



## Modelos de fragilidade compartilhados paramétricos: Modelo

### Modelo

Um painel opcional para especificar as opções e as configurações do modelo

#### Distribuição do tempo de sobrevivência

Uma opção para especificar a distribuição do tempo de sobrevivência.

##### Weibull

Especifica a distribuição de Weibull. Esta é a definição padrão.

##### Exponencial

Especifica a distribuição exponencial.

##### Log-normal

Especifica a distribuição de Log-normal

##### Logística de log

Especifica a distribuição logística de log

#### Configurações de covariável

Especifique variáveis covariáveis.

#### Configurações do fator

Especifique as variáveis de fator

#### Valor inicial de interceptação

Uma opção para especificar o valor inicial do termo de interceptação. Se especificado, deve ser um valor numérico único e não pode ser 0.

#### Valor inicial do parâmetro de escala

Uma opção para controlar a configuração do parâmetro de escala.

#### Erro padrão da regressão OLS correspondente

Usa o erro padrão da regressão de quadrados mínimos ordinários correspondente como o valor inicial.

#### Inverte erro padrão da regressão OLS correspondente

Usa a recíproca do erro padrão

#### Valor fornecido pelo usuário

Se um valor numérico único for especificado, o valor será usado como valor inicial. Se especificado, deve ser maior que 0.

#### Componente de fragilidade

Um parâmetro opcional para especificar a **Distribuição** do componente de fragilidade

##### Gama

Especifica a distribuição gama. Esta é a definição padrão.

##### Inverso-Gaussiano

Especifica a distribuição inversa-gaussiana

#### Valor inicial da variância

Especifica o valor inicial da variação do componente de fragilidade. Deve ser um valor numérico único maior que 0. O valor padrão é 1.0 para distribuição Gama e 0.1 para distribuição Gaussiana inversa.

## Modelos Paramétricos de Fragilidade Compartilhados: Estimativa

### Estimativa

Um painel opcional para especificar as configurações para controlar a estimação dos modelos de fragilidade compartilhados e o processo de seleção de recurso opcional

## Método ou Multiplicadores de Direção Alternados (ADMM)

### Rápido

Aplica o método de direção alternada rápida de multiplicadores (ADMM) Esse é o padrão.

### Tradicional

Aplica o algoritmo tradicional do ADMM

## Aplicar regularização L-1

Conduz o processo para controlar a seleção de variável O campo **Parâmetro de Penalidade** especifica o parâmetro de penalidade que controla o processo de regularização.. Deve ser um valor único maior que 0. A configuração padrão é 0.001..

## Critérios de convergência modelo

### Convergência de parâmetro

Especifica os critérios de convergência para o parâmetro Ele deve ser um valor numérico único pertencente a [0, 1). A configuração padrão é 0.000001. Para **Tipo**, é possível selecionar **ABSOLUTE** para aplicar a convergência absoluta à otimização interna ou **RELATIVE** para aplicar a convergência relativa à otimização interna. O **Valor** opcional especifica um limite numérico para o tipo de convergência.

### Convergência de função objetiva

Especifica os critérios de convergência para a função objetivo. Ele deve ser um valor numérico único que pertence a [0, 1). A configuração padrão é 0, que não aplica os critérios de convergência. Para **Tipo**, é possível selecionar **ABSOLUTE** para aplicar a convergência absoluta à otimização interna ou **RELATIVE** para aplicar a convergência relativa à otimização interna. O **Valor** opcional especifica um limite numérico para o tipo de convergência.

### Convergência Hessiana

Especifica os critérios de convergência para a matriz Hessiana. Ele deve ser um valor numérico único que pertence a [0, 1). A configuração padrão é 0, que não aplica os critérios de convergência. Para **Tipo**, é possível selecionar **ABSOLUTE** para aplicar a convergência absoluta à otimização interna ou **RELATIVE** para aplicar a convergência relativa à otimização interna. O **Valor** opcional especifica um limite numérico para o tipo de convergência.

## Critérios de convergência residual

Uma opção para controlar o processo de otimização.

### Resíduos primários e duplos

Aplica o critério de convergência residual primário e dual. Essa configuração é por padrão.

### Apenas residual primário

Aplica o critério de convergência residual primordial.

### Apenas residual duplo

Aplica o critério de convergência residual dual.

## Método

Um parâmetro opcional para especificar o método de estimação

### Automático

Escolhe automaticamente o método baseado no conjunto de dados de amostra. Esse método é selecionado por padrão.. O campo **Número limite de preditores** especifica o limite do número de preditores e deve ser um número inteiro único maior que 1. O valor padrão é 1000.

### Newton-Raphson

Aplica o método de Newton-Raphson.

### L-BFGS

Aplica o algoritmo de BFGS de memória limitada O campo **Atualizar** especifica o número de atualizações passadas que são mantidas pelo algoritmo BFGS de memória limitada e deve ser um número inteiro maior ou igual a 1. O valor padrão é 5 segundos.

## Iteração

### Máximo de iterações

Especifica o número máximo de iterações. Ele deve ser um número inteiro único que pertence a [1, 300]. A configuração padrão é 20.

### Máxima divisão de passos pela metade

Especifica o número máximo de redução da etapa pela metade. Ele deve ser um número inteiro único que pertence a [1, 200]. A configuração padrão é 5.

### Número máximo de procuras de linha

Especifica o número máximo de procuras de linhas. Ele deve ser um número inteiro único que pertence a [1, 300]. A configuração padrão é 20.

### Convergência absoluta para o processo de iteração

Especifica a convergência absoluta para o processo de iteração externa. Ele deve ser um único valor numérico que pertence a (0, 1). A configuração padrão é 0.0001..

### Convergência relativa para processo de iteração

Especifica a convergência relativa para o processo de iteração externa. Ele deve ser um único valor numérico que pertence a (0, 1). A configuração padrão é 0.01..

## Modelos de fragilidade compartilhados paramétricos: Impressão

### Imprimir

Um painel opcional que controla as saídas da tabela

### Detalhes de codificação do fator

Se selecionado, exibe e imprime os detalhes de codificação dos fatores. O processo será ignorado se não houver fatores em vigor..

### Valores iniciais que são designados aos parâmetros de regressão

Se selecionado, exibe os valores iniciais usados no processo de estimação.

### Histórico de iteração do modelo

Se selecionado, exibe o histórico de iteração da análise de sobrevivência. No campo **Número de etapas**, especifique o número de etapas entre 1 e 99999999. A definição padrão é 1.

## Modelos de fragilidade compartilhados paramétricos: prever

### Prever

Um painel opcional para pontuar e salvar as estatísticas preditas para o conjunto de dados ativo

### Valores de tempo para escoragem

#### Valores de tempo definidos por variáveis dependentes

Pontua as **Predições** com base na variável de tempo especificada para o modelo de sobrevivência paramétrica..

#### Intervalos regulares

Pontua as **Predições** com base em valores de tempo futuros. O campo **Intervalo de Tempo** especifica o intervalo de tempo e deve ser um valor numérico único maior que 0. O campo **Número de períodos de tempo** especifica o número dos períodos de tempo e deve ser um número inteiro numérico único entre 2 e 100.

#### Tempo de duração

Pontua as **Predições** com base na duração de tempo para definir os valores de tempo futuros. Ela deve ser uma única variável numérica.

## Predições

### Sobrevivência

Pontua e salva as estatísticas de sobrevivência previstas no conjunto de dados ativo. O nome da variável customizada padrão (ou nome-raiz) é `PredSurvival`.

### Risco

Pontua e salva os perigos previstos no conjunto de dados ativo. O nome da variável customizada padrão (ou nome-raiz) é `PredHazard`.

### Risco acumulativo

Pontua e salva os perigos acumulativos previstos no conjunto de dados ativo. O nome da variável customizada padrão (ou nome-raiz) é `PredCumHazard`.

### Sobrevivência condicional

Escores e salva as estatísticas de sobrevivência condicional predita para o conjunto de dados ativo. O nome da variável customizada padrão (ou nome-raiz) é `PredConditionalSurvival`. O processo será ignorado se `PASTTIME` não for especificado. Um valor **Tempo de sobrevivência passado** é necessário e especifica os valores de tempo passado para escoragem. Ela deve ser uma única variável numérica.

### Sobrevida incondicional

Escore e salva as estatísticas de sobrevivência incondicional previstas no conjunto de dados ativo. A palavra-chave é suprimida por padrão.. Se especificado, ele poderia ser seguido por um nome de variável opcional fornecido pelo usuário (ou nome raiz) especificado entre parênteses. O nome padrão é `PredUnCondSurvival`..

### Risco incondicional

Escores e salva as estatísticas de risco incondicional predito para o conjunto de dados ativo. A palavra-chave é suprimida por padrão.. Se especificado, ele poderia ser seguido por um nome de variável opcional fornecido pelo usuário (ou nome raiz) especificado entre parênteses. O nome padrão é `PredUncondHazard`..

### Risco de cum incondicional

Escores e salva as estatísticas de risco acumulativo incondicional predito no conjunto de dados ativo. A palavra-chave é suprimida por padrão.. Se especificado, ele poderá ser seguido por um nome de variável opcional fornecido pelo usuário (ou nome raiz especificado entre parênteses. O nome padrão é `PredUncondCumHazard`..

## Modelos de Fragilidade Compartilhados Paramétricos: Plot

### Gráfico

#### Gráficos de função

Uma opção para controlar os gráficos de função

#### Tipo

##### Sobrevivência

Cria o gráfico para as funções de sobrevivência incondicional.

##### Risco

Cria o gráfico para as funções de risco incondicional.

##### Densidade

Cria um gráfico para as funções de densidade..

#### Número de pontos a exibir

Especifica o número de pontos de função entre 1 e 200. A configuração padrão é 100.

#### Valores covariados para gráfico

Uma opção para especificar valores fornecidos pelo usuário e designá-los aos preditores. Por padrão, os gráficos designados serão criados na Média de cada covariável em vigor.. Se especificado, os gráficos designados serão criados com base na configuração do padrão.. Na presença de quaisquer variáveis duplicadas, o especificado primeiro seria reconhecido e o

restante seria ignorado.. Uma variável válida deve estar contida em um efeito de modelo Para uma covariável, o valor fornecido pelo usuário deve ser numérico. A omissão de uma variável em efeito indica que a Média seria usada por padrão para a covariável.. Se um valor inválido for designado a uma variável, o padrão solicitado não será plotado.

#### **Valores de fator para gráfico**

Uma opção para especificar valores fornecidos pelo usuário e designá-los aos preditores. Por padrão, os gráficos designados serão criados na frequência de categoria de cada fator em vigor Se especificado, os gráficos designados serão criados com base na configuração do padrão.. Na presença de quaisquer variáveis duplicadas, o especificado primeiro seria reconhecido e o restante seria ignorado.. Uma variável válida deve estar contida em um efeito de modelo A omissão de uma variável em efeito indica que a frequência da categoria seria usada por padrão para o fator.. Se um valor inválido for designado a uma variável, o padrão solicitado não será plotado.

#### **Linhas separadas para**

Uma opção para especificar uma variável categórica pela qual os gráficos de linha serão desenhados.

#### **Número máximo de linhas em um gráfico**

Especifica o número máximo das linhas em um gráfico se **Separar linhas para** for especificado A configuração padrão é 10.

## **Modelos paramétricos compartilhados de fragilidade: Exportação**

### **Exportar**

Selecione **Exportar informações de modelo para arquivo XML** para gravar as informações de modelo e parâmetro em um arquivo PMML para escoragem. Você deve especificar o diretório e o nome do arquivo PMML a ser salvo.

## **Modelos de fragilidade compartilhados paramétricos: definir eventos**

Uma opção para definir status. Se a variável de status for omitida, falha ou evento se tornará o status padrão para todos os casos.

1. A partir do menu escolher,

**Analisar > Sobrevivência > Modelos Paramétricos Compartilhados de fragilidade ...**

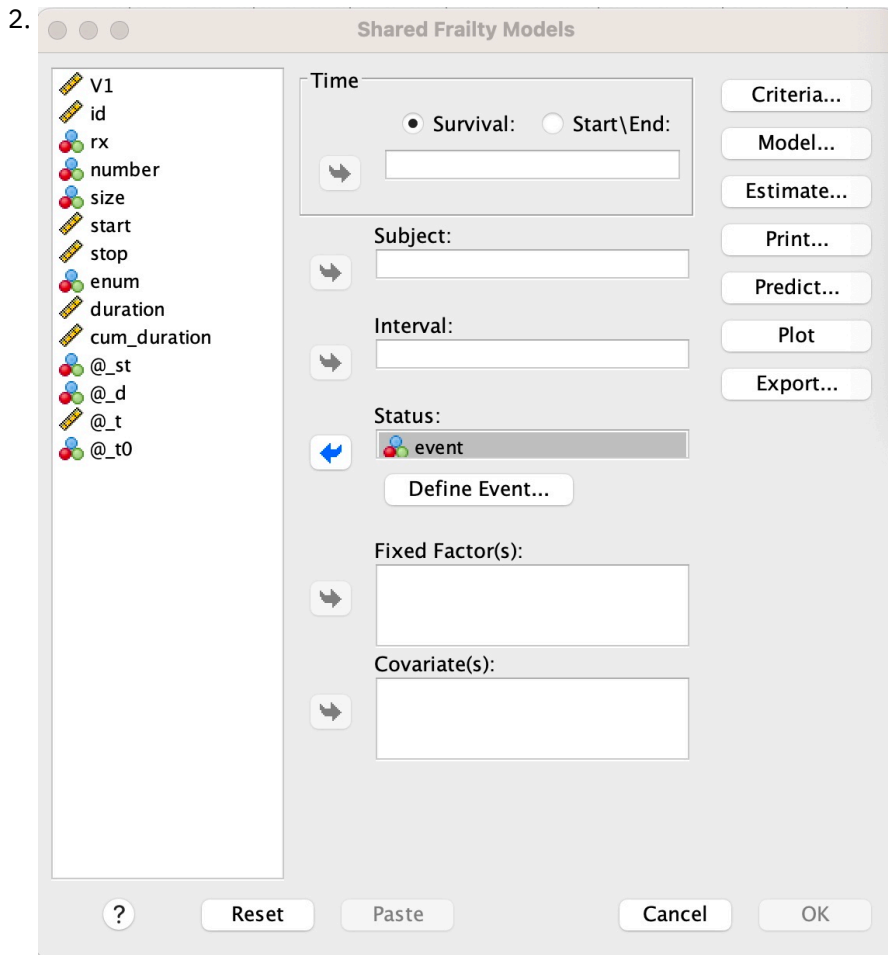


Figura 5. Modelos de fragilidade compartilhados-caixa de diálogo-Status

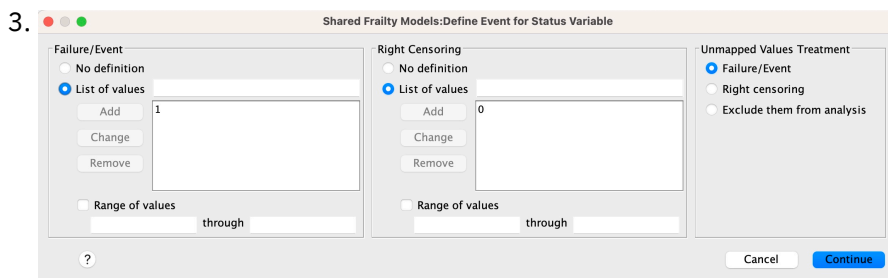


Figura 6. Modelos de fragilidade compartilhados-caixa de diálogo-Status-Definir evento

## Modelos de fragilidade compartilhados paramétricos-Exemplos

### Exemplo 1

SURVREG RECORRENTE y COM x1 BY x2

/MODEL SUBJECT = id FRAILTY=GAMA DISTRIBUTION=WEIBULL.

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de y em uma covariável x1 e no fator x2.

O tempo de sobrevivência é representado por uma única variável y.

Os sujeitos são identificados pela variável id.

O tempo de sobrevivência é assumido para seguir uma distribuição Weibull.

Assume-se que a variância da fragilidade segue uma distribuição Gama.

Todos os registros válidos são usados na análise de sobrevida.

### Exemplo 2

```
SURVREG RECORRENTE y COM x1 BY x2
```

```
/MODEL SUBJECT = id FRAILTY=INV_GAUSSIAN DISTRIBUTION=LOG_NORMAL INTERVAL=z.
```

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de y em uma covariável x1 e no fator x2.

O tempo de sobrevivência é representado por duas variáveis y1 e y2 denotando o horário de início e de encerramento.

Os sujeitos são identificados pela variável id.

Presume-se que o tempo de sobrevivência siga uma distribuição Log-normal

Assume-se que a variância da fragilidade segue uma distribuição Gaussiana inversa.

Intervalos de tempo são definidos pela variável z. Para cada sujeito, o procedimento usa apenas os registros não conflitantes e exclui da análise todos os registros após o primeiro status de falha

### Exemplo 3

```
SURVREG RECORRENTE y1 y2 WITH x1 BY x2(1)
```

```
/MODEL SUBJECT = id FRAILTY=INV_GAUSSIAN DISTRIBUTION=LOG_LOGISTIC
```

```
/STATUS VARIABLE=evento FAILURE=1 RIGHT=0.
```

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de y em uma covariável x1 e no fator x2. Sobrevivência

é representado por duas variáveis y1 e y2 denotando o horário de início e de encerramento, respectivamente. Para o fator x2, a categoria "1" é designada como uma linha de base a ser modelada.

Os sujeitos são identificados pela variável id.

O tempo de sobrevivência é assumido para seguir uma distribuição Log-logística.

Assume-se que a variância da fragilidade segue uma distribuição Gaussiana inversa.

O evento de variável é especificado para definir o status com 1 e 0 denotando falha e censura direita, respectivamente.

### Exemplo 4

```
SURVREG RECORRENTE y COM x1 BY x2
```

```
/ASSUNTO DO MODELO = id
```

```
/STATUS VARIABLE=evento FAILURE=1 RIGHT=0
```

```
/PREVER UNCONDSURVIVAL UNCONDHAZARD UNCONDCUMHAZARD
```

```
/FUNCTIONPLOT SURVIVAL HAZARD DENSITY PLOTBY (x2).
```

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de y em uma covariável x1 e no fator x2. O tempo de sobrevivência é representado por uma única variável y.

Os sujeitos são identificados pela variável id.

Sobrevivência incondicional ou baseada em população, risco e risco acumulativo são pontuados e salvos no conjunto de dados ativo.

As curvas de sobrevivência e de risco incondicionais ou baseadas em população são plotadas separadas pelas categorias em x2.

### Exemplo 5

SURVREG RECORRENTE y COM x1 BY x2

/MODEL SUBJECT = id FRAILITY=GAMMA DISTRIBUTION=WEIBULL

/STATUS VARIABLE=evento FAILURE=1 RIGHT=0

/ESTIMATION HCONVERGE=1e-12(RELATIVE ) PCONVERGE=0 FCONVERGE=0 SELECT FEATURES=TRUE PENALTY=0.01.

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de y em uma covariável x1 e no fator x2. O tempo de sobrevivência é representado por uma única variável y.

Os sujeitos são identificados pela variável id.

O tempo de sobrevivência é assumido para seguir uma distribuição Weibull.

Assume-se que a variância da fragilidade segue uma distribuição Gama.

Os critérios de convergência são baseados na matriz Hessiana. Ele usa 1e-12 como convergência relativa.

O modelo inclui um termo de penalidade para controlar o processo de regularização O parâmetro de multa é configurado como 0.01.

### Exemplo 6

SURVREG RECORRENTE y COM x1 BY x2

/ASSUNTO DO MODELO = id

/STATUS VARIABLE=infetar FAILURE=1 RIGHT=0

/ESTIMATION MAXLINESEARCH=100 MAXITER=50 MAXSTEPHALVING=20

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de y em uma covariável x1 e no fator x2. O tempo de sobrevivência é representado por uma única variável y.

Os sujeitos são identificados pela variável id.

O procedimento especifica o número máximo da procura de linha para ser 100, o número máximo de iterações para ser 50 e o número máximo de redução da etapa para ser 20

## Modelos paramétricos compartilhados de fragilidade-um estudo de caso para dados recuperados

Modelos paramétricos compartilhados de fragilidade-um estudo de caso para dados recuperados

Nome do caso de uso-Efeito colateral do tratamento.

Atores-Investigador e profissional de saúde pública.

Pré-condições-Um conjunto de dados limpo disponível com base no tempo de sobrevivência, status do efeito secundário e preditores a serem ajustados.

Descrição-Patrick, investigador de saúde pública, está investigando uma amostra de dados que inclui 20 participantes. Estes participantes são recrutadas num estudo sobre um efeito secundário ligeiro que é potencialmente causado por um novo tratamento. O designer de tratamento afirma que não haveria diferenças entre homens e mulheres, em relação ao efeito colateral. Patrick gostaria de avaliar tal hipótese. As variáveis incluídas na amostra de dados são listadas da seguinte forma:

- patID: Número de identificação para identificar um único participante.



- endTime: Tempo de sobrevivência (em dias) do efeito colateral após um tratamento, medido desde o início do tratamento até o relato de um efeito colateral ou a censura em 60 dias.
- sideEffect: Status do efeito colateral, status = 0 se censurado e status = 1 se o efeito colateral leve for relatado.
- idade: idade do participante no período da pesquisa.
- fêmea: fêmea = 0 se macho e fêmea = 1 se fêmea.

Diversos tratamentos podem ser aplicados, o que resulta em diversos registros de tempos de recorrência medidos para um determinado participante. O horário de início é sempre 0 para cada registro, que é omitido na amostra de dados. Patrick está interessado em visualizar as funções de sobrevivência e risco para desenhar uma comparação entre um macho e uma fêmea, controlando sua idade e fragilidade. Ele está ciente de que os tratamentos que são administrados ao mesmo participante são mais correlacionados. Ao assumir que o tempo de sobrevivência segue uma distribuição Weibull, Patrick decide construir um modelo paramétrico de sobrevivência de fragilidade compartilhada no SPSS Statistics para considerar a dependência de tratamento para o mesmo participante.

Sintaxe-

```
DATA LIST FREE
/patID(F5.0) endTime(F5.0) sideEffect(F2.0) age(F5.2) female(F2.0) .
BEGIN DATA .
1 45 0 38.00 0
2 26 1 20.00 1
3 58 0 53.00 0
4 31 1 37.00 1
4 24 0 37.00 1
4 50 0 37.00 1
5 20 1 51.00 0
5 38 1 51.00 0
6 30 0 35.00 1
7 22 1 58.00 1
8 53 1 29.00 1
8 49 1 29.00 1
9 25 0 45.00 0
9 25 0 45.00 0
10 27 0 33.00 1
11 34 1 21.00 1
11 40 0 21.00 1
11 49 0 21.00 1
12 42 1 26.00 0
13 25 0 40.00 0
14 21 1 52.00 0
14 32 1 52.00 0
15 56 0 28.00 1
15 34 0 28.00 1
16 30 0 41.00 0
16 29 0 41.00 0
17 25 1 27.00 0
18 26 1 54.00 1
18 36 1 54.00 1
19 27 0 39.00 0
20 58 1 22.00 1
20 54 0 22.00 1
20 43 1 22.00 1
END DATA.
SURVREG RECURRENT endTime WITH age BY female
/MODEL SUBJECT=patID FRAILTY=GAMMA DISTRIBUTION=WEIBULL
/ESTIMATION HCONVERGE=1e-12 PCONVERGE=0 FCONVERGE=0
/STATUS VARIABLE=sideEffect FAILURE=1 RIGHT=0
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY(female) .
```

### Sinopse:

A sintaxe especificada por Patrick designa endTime como uma única variável de tempo dependente. O procedimento assume automaticamente que o horário de início é 0 para cada registro.. As variáveis idade e sexo feminino são modeladas como covariável e fator, respectivamente. Supõe-se que os tempos de sobrevivência de recorrência seguem uma distribuição Weibull. Presume-se que o termo fragilidade não observado segue uma distribuição Gamma e seu componente de variância é modelado. Em relação às saídas, a tabela Resumo do Modelo fornece as informações de procedimento e modelo. A tabela Resumo

de Processamento de Caso fornece uma listagem abrangente do status de falha / censura e também os casos que são excluídos da análise

Na amostra dos dados de Patrick, todos os registros são válidos e incluídos na análise. Ao comparar o log da verossimilhança com aquele do modelo correspondente sem o componente de fragilidade, o modelo de fragilidade compartilhada falha ao atingir um nível significativo (valor  $p = 0.168$ ). Patrick está se perguntando se é necessário incluir um termo de fragilidade compartilhada no modelo. O fator de aceleração estimado de um participante masculino é 1.017, que é obtido calculando o expoente do coeficiente de regressão estimado 0.017 de [feminino = 0.0]. Seu intervalo de confiança de 95% associado é (.688, 1.504) Esses resultados sugerem que um indivíduo do sexo masculino tem quase o mesmo fator de aceleração que um indivíduo do sexo feminino com a mesma idade e fragilidade. No nível de população, Patrick plota as curvas de sobrevivência incondicional e de risco separadamente para homens e mulheres que são avaliados na média de idade da amostra (37.45 anos).

Patrick confirma que, para qualquer valor fixo de tempo de sobrevivência, um homem e uma mulher, em média, devem ter a mesma probabilidade de sobrevivência. Curiosamente, apesar de uma forma unimodal que é mostrada no gráfico de risco incondicional, Patrick descobre que dentro de um período de 60 dias o risco da população está realmente aumentando. Esse comportamento pode implicar a existência do efeito fragilidade. Para investigar melhor o efeito colateral causado pelos tratamentos, Patrick pode continuar com um modelo sem o componente de fragilidade e comparar o comportamento de homens e mulheres. Além disso, ele pode considerar o acompanhamento com os participantes por um período superior a 60 dias para coletar mais dados.

## Análise de Sobrevivência de Kaplan-Meier

---

Há muitas situações nas quais você gostaria de examinar a distribuição de vezes entre dois eventos, como duração de emprego (tempo entre ser contratado e deixar a empresa). No entanto, esse tipo de dados geralmente inclui alguns casos censurados. Casos censurados são casos para os quais o segundo evento não está registrado (por exemplo, pessoas que ainda trabalham para a empresa no final do estudo). O procedimento de Kaplan-Meier é um método de estimação de modelos de tempo para evento na presença de casos censurados. O modelo de Kaplan-Meier é baseado na estimação de probabilidades condicionais em cada ponto de tempo em que um evento ocorre e tomando o limite do produto daquelas probabilidades para estimar a taxa de sobrevivência em cada ponto no tempo.

**Exemplo.** Um novo tratamento para AIDS tem benefício terapêutico em estender a vida? Você poderia realizar um estudo usando dois grupos de pacientes de AIDS, um recebendo terapia tradicional e o outro recebendo o tratamento experimental. Construir um modelo de Kaplan-Meier a partir dos dados lhe permitiria comparar taxas de abstinência globais entre os dois grupos para determinar se o tratamento experimental é uma melhoria em relação à terapia tradicional. Também é possível criar gráfico de funções de sobrevivência ou de risco e compará-las visualmente para obter informações mais detalhadas.

**Estatísticas.** Tabela de sobrevivência, incluindo tempo, status, sobrevivência acumulativa e erro padrão, eventos acumulativos e número restante; e tempo de sobrevivência de média e mediana, com erro padrão e 95% de intervalo de confiança. Gráficos: sobrevivência, risco, log de sobrevivência e um menos a sobrevivência.

Considerações de Dados de Kaplan-Meier

**Dados.** A variável de horário deve ser contínua, a variável de status pode ser categórica ou contínua e as variáveis de fator e estrato devem ser categóricas.

**Suposições.** Probabilidades para o evento de interesse devem depender apenas do tempo após o evento inicial--elas são consideradas como estáveis com relação ao tempo absoluto. Isso é, casos que entram no estudo em tempos diferentes (por exemplo, pacientes que começaram o tratamento em tempos diferentes) devem se comportar de forma semelhante. Também não deve haver diferenças sistemáticas entre casos censurados e não censurados. Se, por exemplo, muitos dos casos censurados forem pacientes com condições mais sérias, os seus resultados podem ser influenciados.

**Procedimentos relacionados.** O procedimento de Kaplan-Meier usa um método para calcular tabelas de mortalidade que estima a função de sobrevivência ou de risco no momento de cada evento. O procedimento de Tabelas de Mortalidade usa uma abordagem atuarial para análise de sobrevivência

que depende de particionar o período de observação em intervalos de tempo menores e pode ser útil para tratar amostras grandes. Se você tiver variáveis que suspeita estarem relacionadas a tempo de sobrevivência ou variáveis que você deseja controlar (covariáveis), use o procedimento de Regressão de Cox. Se as suas covariáveis podem ter valores diferentes em pontos diferentes no tempo para o mesmo caso, use Regressão de Cox com Covariáveis Dependentes de Tempo.

Obtendo uma Análise de Sobrevivência de Kaplan-Meier

1. Nos menus, escolha:

**Analisar > Sobrevivência > Kaplan-Meier...**

2. Selecione uma variável de tempo.

3. Selecione uma variável de status para identificar casos para os quais o evento terminal ocorreu. Essa variável pode ser numérica ou *sequência curta*. Em seguida, clique em **Definir Evento**.

Opcionalmente, é possível selecionar uma variável de fator para examinar diferenças de grupo. Também é possível selecionar uma variável de estrato, que produzirá análises separadas para cada nível (estrato) da variável.

## Kaplan-Meier - Definir Evento para Variável de Status

Insira o valor ou os valores indicando que o evento terminal ocorreu. É possível inserir um valor único, um intervalo de valores ou uma lista de valores. A opção de Intervalo de Valores está disponível somente se a sua variável de status for numérica.

## Kaplan-Meier - Comparar Níveis do Fator

É possível solicitar estatísticas para testar a igualdade das distribuições de sobrevivência para os níveis diferentes do fator. Estatísticas disponíveis são ranqueamento de log, Breslow e Tarone-Ware. Selecione uma das alternativas para especificar as comparações a serem feitas: agrupadas no estrato, para cada estrato, pairwise no estrato ou pairwise para cada estrato.

**Tendência linear para níveis do fator.** Permite que você teste para uma tendência linear nos níveis do fator. Essa opção está disponível somente para comparações gerais (em vez de pairwise) de níveis do fator.

- *Classificação de log.* Um teste para comparar a igualdade de distribuições de sobrevivência.. Todos os pontos de tempo são ponderados igualmente neste teste.
- *Breslow.* Um teste para comparar a igualdade de distribuições de sobrevivência.. Os momentos são ponderados pelo número de casos de risco em cada momento.
- *Tarone-Ware.* Um teste para comparar a igualdade de distribuições de sobrevivência.. Os momentos são ponderados pela raiz quadrada do número de casos sob risco em cada momento.
- *Agrupadas no estrato.* Compara todos os níveis de fatores em um único teste para testar a igualdade das curvas de sobrevivência.
- *Pairwise no estrato.* Compara cada par distinto de níveis de fator. Os testes de tendência entre pares não estão disponíveis.
- *Para cada estrato.* Executa um teste separado de igualdade de todos os níveis de fator para cada estrato. Se você não tiver uma variável de estratificação, os testes não serão executados.
- *Pairwise para cada estrato.* Compara cada par distinto de níveis de fator para cada camada.. Os testes de tendência entre pares não estão disponíveis. Se você não tiver uma variável de estratificação, os testes não serão executados.

## Kaplan-Meier - Salvar Novas Variáveis

É possível salvar informações a partir de sua tabela de Kaplan-Meier como novas variáveis, que podem, então, ser usadas em análises subsequentes para testar hipóteses ou verificar premissas. É possível salvar eventos de sobrevivência, de erro padrão de sobrevivência, risco e acumulativos como novas variáveis.

- *Sobrevivência*. Estimativa de probabilidade de sobrevivência cumulativa. O nome da variável padrão é o prefixo `sur_` com um número sequencial anexado a ele. Por exemplo, se `sur_1` já existir, o Kaplan-Meier designará o nome de variável `sur_2`.
- *Erro padrão de sobrevivência*. Erro padrão da estimativa de sobrevivência acumulativa. O nome da variável padrão é o prefixo `se_` com um número sequencial anexado a ele. Por exemplo, se `se_1` já existir, o Kaplan-Meier designará o nome de variável `se_2`.
- *Risco*. Estimativa de função de risco cumulativo. O nome da variável padrão é o prefixo `haz_` com um número sequencial anexado a ele. Por exemplo, se `haz_1` já existir, o Kaplan-Meier designará o nome de variável `haz_2`.
- *Eventos acumulativos*. Frequência cumulativa de eventos quando os casos são classificados por seus tempos de sobrevivência e códigos de status. O nome da variável padrão é o prefixo `cum_` com um número sequencial anexado a ele. Por exemplo, se `cum_1` já existir, o Kaplan-Meier designará o nome de variável `cum_2`.

## Opções de Kaplan-Meier

É possível solicitar vários tipos de saída a partir da análise de Kaplan-Meier.

**Estatísticas.** É possível selecionar estatísticas exibidas para as funções de sobrevivência calculadas, incluindo tabela(s) de sobrevivência, sobrevivência média e mediana e quartis. Se você tiver incluído variáveis de fator, estatísticas separadas serão geradas para cada grupo.

**Gráficos.** Gráficos lhe permitem examinar as funções de sobrevivência, de sobrevivência um menos, de risco e de log de sobrevivência visualmente. Se você tiver incluído variáveis de fator, funções serão criadas em gráfico para cada grupo.

- *Sobrevivência*. Exibe a função de sobrevivência acumulativa em uma escala linear.
- *Um menos sobrevivência*. Exibe a função um menos a sobrevivência em uma escala linear.
- *Risco*. Exibe a função de risco acumulativo em uma escala linear.
- *Sobrevivência de log*. Exibe a função de sobrevivência acumulativa em uma escala logarítmica.

## Variáveis Adicionais de Comando do KM

O idioma da sintaxe de comando também permite:

- Obter tabelas de frequências que considerem casos perdidos para acompanhamento como uma categoria separada de casos censurados.
- Especificar espaçamento desigual para o teste para tendência linear.
- Obter percentis diferentes de quartis para a variável de tempo de sobrevivência.

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

## Análise de Regressão de Cox

A regressão de Cox constrói um modelo preditivo para dados de tempo para evento. O modelo produz uma função de sobrevivência que prevê a probabilidade de que o evento de interesse tenha ocorrido em um determinado momento  $t$  para determinados valores das variáveis preditoras. A forma da função de sobrevivência e os coeficientes de regressão para os preditores são estimados a partir de assuntos observados; o modelo pode então ser aplicado em novos casos que têm medidas para as variáveis preditoras. Observe que informações de assuntos censurados, ou seja, aqueles que não experimentam o evento de interesse durante o tempo de observação, contribuem de modo útil para a estimação do modelo.

**Exemplo,** Homens e mulheres têm riscos diferentes de desenvolver câncer de pulmão com base no consumo de cigarros? Construindo um modelo de Regressão de Cox, com o uso de cigarros (cigarros fumados por dia) e gênero inseridos como covariáveis, é possível testar hipóteses relativas aos efeitos de gênero e uso de cigarros no tempo para o início de câncer de pulmão.

**Estatísticas.** Para cada modelo:  $-2LL$ , a estatística de razão de verossimilhança e o qui-quadrado geral. Para variáveis no modelo: estimativas de parâmetros, erros padrão e estatísticas de Wald. Para variáveis não no modelo: estatísticas de escore e qui-quadrado de resíduo.

Considerações de Dados de Regressão de Cox

**Dados.** A sua variável de horário deve ser quantitativa, mas a sua variável de status pode ser categórica ou contínua. Variáveis independentes (covariáveis) podem ser contínuas ou categóricas; se categóricas, elas devem ser codificadas por simulado ou por indicador (há uma opção no procedimento para recodificar variáveis categóricas automaticamente). Variáveis de camada devem ser categóricas, codificadas como números inteiros ou sequências curtas.

**Suposições.** Observações devem ser independentes e a razão de risco deve ser constante ao longo do tempo; isto é, a proporcionalidade de riscos de um caso para outro não deve variar ao longo do tempo. A última suposição é conhecida como a **suposição de riscos proporcionais**.

**Procedimentos relacionados.** Se a suposição de riscos proporcionais não se mantiver (veja acima), você pode precisar usar o procedimento de Covariáveis Dependentes de Tempo de Cox. Se você não tiver covariáveis ou se tiver somente uma covariável categórica, é possível usar o procedimento de Tabelas de Mortalidade ou Kaplan-Meier para examinar funções de sobrevivência ou risco para a(s) sua(s) amostra(s). Se você não tiver dados censurados em sua amostra (isto é, todos os casos experimentaram o evento terminal), é possível usar procedimento de Regressão Linear para modelar o relacionamento entre preditores e tempo para evento.

Obtendo uma Análise de Regressão de Cox

1. Nos menus, escolha:

**Analisar > Sobrevivência > Regressão de Cox...**

2. Selecione uma variável de tempo. Casos cujos valores de tempo não são analisados.

3. Selecione uma variável de status e, em seguida, clique em **Definir Evento**.

4. Selecione uma ou mais covariáveis. Para incluir termos de interação, selecione todas as variáveis envolvidas na interação e, em seguida, clique em **> a\*b>**.

Como opção, é possível calcular modelos separado para grupos diferentes definindo uma variável de camada.

## Regressão de Cox - Definir Variáveis Categóricas

É possível especificar detalhes de como o procedimento de Regressão de Cox tratará variáveis categóricas.

**Covariáveis.** Lista todas as covariáveis especificadas na caixa de diálogo principal, por si mesmas ou como parte de uma interação, em qualquer camada. Se algumas dessas forem variáveis de sequência de caracteres ou forem categóricas, será possível usá-las somente como covariáveis categóricas.

**Covariáveis categóricas.** Lista variáveis identificadas como categóricas. Cada variável inclui uma notação entre parênteses indicando a codificação de contraste a ser usada. As variáveis de sequência de caracteres (denotadas pelo símbolo < após seus nomes) já estão presentes na lista Covariáveis categóricas. Selecione quaisquer outras covariáveis categóricas da lista Covariáveis e mova-as para a lista Covariáveis categóricas.

**Mudar contraste.** Permite mudar o método de contraste. Os métodos de contraste disponíveis são:

- **Indicador.** Os contrastes indicam a presença ou a ausência de associação de categoria. A categoria de referência é representada na matriz de contraste como uma linha de zeros.
- **Simples.** Cada categoria da variável preditora, exceto a categoria de referência, é comparada com a categoria de referência.
- **Diferença.** Cada categoria da variável preditora, exceto a primeira categoria, é comparada com o efeito médio de categorias anteriores. Também conhecido como contrastes de Helmert reversos.
- **Helmert.** Cada categoria da variável preditora, exceto a última categoria, é comparada com o efeito médio de categorias subsequentes.

- **Repetido.** Cada categoria da variável preditora, exceto a primeira categoria, é comparada com a categoria que a precede.
- **Polinomial.** Contrastes polinomiais ortogonais. As categorias são consideradas igualmente espaçadas. Os contrastes polinomiais estão disponíveis apenas para variáveis numéricas.
- **Desvio.** Cada categoria da variável preditora, exceto a categoria de referência, é comparada com o efeito geral.

Se você selecionar **Desvio**, **Simples** ou **Indicador**, selecione **Primeiro** ou **Último** como a categoria de referência. Observe que o método não será realmente mudado até você clicar em **Mudar**.

As covariáveis de sequência de caracteres devem ser covariáveis categóricas. Para remover uma variável de sequência de caracteres da lista Covariáveis categóricas, deve-se remover todos os termos contendo a variável da lista Covariáveis na caixa de diálogo principal.

## Gráficos de Regressão de Cox

Os gráficos podem ajudar a avaliar seu modelo estimado e a interpretar os resultados. É possível representar as funções de sobrevivência, de risco, de log menos log e de um menos sobrevivência.

### Tipo de gráfico

#### Sobrevivência

Exibe a função de sobrevivência acumulativa em uma escala linear.

#### Risco

Exibe a função de risco acumulativo em uma escala linear.

#### Log menos log

A estimativa de sobrevivência acumulativa após a transformação  $\ln(-\ln)$  é aplicada à estimativa.

#### Um menos sobrevivência

Exibe a função um menos a sobrevivência em uma escala linear.

### Valores covariáveis plotados em

É possível plotar uma linha separada para cada valor de covariável categórica, movendo essa covariável para a caixa de texto **Separar linhas para**. Esta opção está disponível somente para covariáveis categóricas, que são indicadas por **(Cat)** após seus nomes na lista **Valores covariáveis plotados em**.

### Alterar Valor

Como essas funções dependem de valores das covariáveis, deve-se usar valores constantes para as covariáveis para criar gráfico das funções versus tempo. O padrão é usar a média de cada covariável como um valor constante, mas é possível inserir seus próprios valores para o gráfico usando o grupo de controle **Mudar valor**.

## Regressão de Cox - Salvar Novas Variáveis

É possível salvar vários resultados de sua análise como novas variáveis. Essas variáveis podem, então, ser usadas em análises subsequentes para testar hipóteses ou verificar premissas.

### Salvar variáveis de modelo

Permite salvar a função de sobrevivência e seu erro padrão, estimativas de log menos log, função de risco, resíduos parciais, DfBeta(s) para a regressão e o preditor linear  $X \cdot \text{Beta}$  como novas variáveis.

#### Função de sobrevivência

O valor da função de sobrevida acumulativa para um determinado tempo. Ele é igual à probabilidade de sobrevivência para esse período de tempo.

#### Erro padrão de função de sobrevivência

Erro padrão da estimativa de sobrevivência acumulativa.

#### Função de sobrevivência de log menos log

A estimativa de sobrevivência acumulativa após a transformação  $\ln(-\ln)$  ser aplicada à estimativa.

**Função de risco**

Salva a estimativa de função de risco acumulativo (também chamado de resíduo de Cox-Snell).

**Resíduos parciais**

É possível plotar resíduos parciais com relação ao tempo de sobrevivência para testar a suposição de riscos proporcionais. Uma variável é salva para cada covariável no modelo final. Os Resíduos Parciais estão disponíveis apenas para os modelos que contiverem pelo menos uma covariável.

**DfBeta(s)**

Mudança estimada em um coeficiente se um caso é removido. Uma variável é salva para cada covariável no modelo final. Os DfBetas estão disponíveis apenas para os modelos que contiverem pelo menos uma covariável.

**X\*Beta**

Escore de preditor linear. A soma do produto dos valores de covariáveis centralizadas na média e suas estimativas paramétrica correspondentes para cada caso.

**Nota:** Somente os DfBetas serão salvos se você estiver executando o Cox com uma covariável dependente de tempo.

**Exportar informações do modelo para o arquivo XML**

As estimativas de parâmetros são exportadas para o arquivo especificado em formato XML. É possível usar esse arquivo de modelo para aplicar as informações de modelo a outros arquivos de dados para propósitos de escoragem.

## Opções de Regressão de Cox

É possível controlar vários aspectos de sua análise e resultado.

**Estatísticas de Modelo.** É possível obter estatísticas para seus parâmetros de modelo, incluindo intervalos de confiança para  $\exp(B)$  e correlação de estimativas. É possível solicitar essas estatísticas em cada passo ou apenas no último passo.

**Probabilidade para stepwise.** Se você tiver selecionado um método stepwise, é possível especificar a probabilidade para a entrada ou a remoção a partir do modelo. Uma variável será inserida se o nível de significância de seu valor  $F$ -para-inserir for menor que o valor de Entrada e uma variável será removida se o nível de significância for maior que o valor de Remoção. O valor de Entrada deve ser menor que o valor de Remoção.

**Máximo de Iterações.** Permite especificar o máximo de iterações para o modelo, que controla por quanto tempo o processo procurará uma solução.

**Exibir função de linha de base.** Permite exibir a função de risco da linha de base e a sobrevivência acumulativa na média das covariáveis. Essa exibição não estará disponível se você tiver especificado covariáveis dependentes de tempo.

## Regressão de Cox - Definir Evento para Variável de Status

Insira o valor ou os valores indicando que o evento terminal ocorreu. É possível inserir um valor único, um intervalo de valores ou uma lista de valores. A opção de Intervalo de Valores está disponível somente se a sua variável de status for numérica.

## Variáveis Adicionais de Comando do COXREG

O idioma da sintaxe de comando também permite:

- Obter tabelas de frequências que considerem casos perdidos para acompanhamento como uma categoria separada de casos censurados.
- Selecionar uma categoria de referência, diferente da primeira ou última, para o desvio, métodos de contraste e de indicador.
- Especificar espaçamento desigual de categorias para o método de contraste polinomial.
- Especificar critérios de iteração adicionais.

- Controlar o tratamento de valores omissos.
- Especificar os nomes para variáveis salvas.
- Gravar resultado em um arquivo de dados externo do IBM SPSS Estatísticas .
- Reter dados para cada grupo de arquivos divididos em um arquivo temporário externo durante o processamento. Isso pode ajudar a conservar recursos de memória ao executar análises com grandes conjuntos de dados. Isso não está disponível com covariáveis dependentes de tempo.

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

## Calculando Covariáveis Dependentes de Tempo

Em determinadas situações, você gostaria de calcular um modelo de regressão Cox, mas a suposição de riscos proporcionais não é válida. Ou seja, as razões de risco mudam ao longo do tempo; os valores de uma (ou mais) de suas covariáveis são diferentes em diferentes pontos de tempo. Nesses casos, é necessário usar um modelo de regressão Cox estendido, no qual você pode especificar **covariáveis dependentes do tempo**.

Para analisar tal modelo, deve-se primeiro definir a covariável dependente de tempo. Para facilitar isso, uma "variável do sistema" que representa tempo está disponível. Essa variável é chamada  $T_{-}$ . É possível usar esta variável para definir covariáveis dependentes de tempo de duas maneiras gerais:

- Para testar a suposição de riscos proporcionais com relação a uma determinada covariável ou estimar um modelo de regressão Cox estendido que permita riscos não proporcionais, é necessário definir a covariável dependente do tempo como uma função da variável de tempo  $T_{-}$  e da covariável em questão. Um exemplo comum seria o produto simples da variável de tempo e a covariável, mas funções mais complexas também podem ser especificadas. Testar a significância do coeficiente da covariável dependente de tempo indica se a suposição de riscos proporcionais é razoável.
- Algumas variáveis podem ter valores diferentes em períodos de tempo diferentes, mas não são sistematicamente relacionadas ao tempo... Nesses casos, é necessário definir uma **covariável dependente de tempo segmentado**, o que pode ser feito usando **expressões lógicas**. Expressões lógicas utilizam o valor 1 se verdadeiro e 0 se falso. Usando uma série de expressões lógicas, é possível criar sua covariável dependente de tempo a partir de um conjunto de medições. Por exemplo, se você tiver pressão arterial que é medida uma vez por semana para as quatro semanas de seu estudo (identificado como  $BP1$  para  $BP4$ ), poderá definir a covariável dependente de tempo como  $(T_{-} < 1) * BP1 + (T_{-} \geq 1 \ \& \ T_{-} < 2) * BP2 + (T_{-} \geq 2 \ \& \ T_{-} < 3) * BP3 + (T_{-} \geq 3 \ \& \ T_{-} < 4) * BP4$ . Observe que exatamente um dos termos entre parênteses é igual a 1 para qualquer caso específico e o restante será igual a 0. Resumindo, essa função significa que se o tempo for menor que uma semana, use  $BP1$ ; se for maior que uma semana, mas menor que duas semanas, use  $BP2$  e assim por diante.

Na caixa de diálogo Covariável dependente de tempo de cálculo, é possível usar os controles de construção de função para construir a expressão para uma covariável dependente de tempo ou é possível inseri-la diretamente na área de texto Nome. Observe que as constantes de sequência devem estar entre aspas ou apóstrofes, e as constantes numéricas devem ser digitadas em formato americano, com o ponto como o delimitador decimal. Quaisquer covariáveis dependentes do tempo resultantes precisam ser incluídas como covariáveis em seu modelo de regressão Cox.

## Calculando uma Covariável Dependente de Tempo

1. No menu, escolha

**Analisar > Sobrevivência > Cox com covariável dependente de tempo...**

2. Insira uma expressão para a covariável dependente de tempo.

3. Clique em **Modelo** para continuar com a Regressão de Cox.

**Observação** - Certifique-se de incluir as novas variáveis adicionadas como covariáveis em seu modelo de regressão Cox.



## Esquemas de codificação de variável categórica

Em muitos procedimentos, é possível solicitar a substituição automática de uma variável independente categórica com um conjunto de variáveis de contraste, que serão, então, inseridos ou removidos de uma equação como um bloco. É possível especificar como o conjunto de variáveis de contraste deve ser codificado, geralmente, no subcomando CONTRAST. Este apêndice explica e ilustra quão diferentes os tipos de contraste solicitados em CONTRAST realmente funcionam.

### Desvio

**Desvio da média global.** Em termos de matriz, esses contrastes têm o formato:

```
mean ( 1/k 1/k ... 1/k 1/k)
df(1) ( 1-1/k -1/k ... -1/k -1/k)
df(2) ( -1/k 1-1/k ... -1/k -1/k)
.
.
df(k-1) ( -1/k -1/k ... 1-1/k -1/k)
```

em que  $k$  é o número de categorias para a variável independente e a última categoria é omitida por padrão. Por exemplo, os contrastes de desvio para uma variável independente com três categorias são as seguintes:

```
( 1/3 1/3 1/3)
( 2/3 -1/3 -1/3)
(-1/3 2/3 -1/3)
```

Para omitir uma categoria diferente da última, especifique o número da categoria omitida entre parênteses após a palavra-chave DEVIATION. Por exemplo, o subcomando a seguir obtém os desvios para as categorias primeira e terceira e omite a segunda:

```
/CONTRAST(FACTOR)=DEVIATION(2)
```

Suponha que *fator* tem três categorias. A matriz de contraste resultante será

```
( 1/3 1/3 1/3)
( 2/3 -1/3 -1/3)
(-1/3 -1/3 2/3)
```

### Simples

**Contrastes simples.** Compara cada nível de um fator com o último. O formulário matriz geral é

```
mean (1/k 1/k ... 1/k 1/k)
df(1) ( 1 0 ... 0 -1)
df(2) ( 0 1 ... 0 -1)
.
.
df(k-1) ( 0 0 ... 1 -1)
```

em que  $k$  é o número de categorias para a variável independente. Por exemplo, os contrastes simples para uma variável independente com quatro categorias são as seguintes:

```
(1/4 1/4 1/4 1/4)
( 1 0 0 -1)
( 0 1 0 -1)
( 0 0 1 -1)
```

Para usar outra categoria em vez da última como uma categoria de referência, especifique entre parênteses após a palavra-chave SIMPLE o número de sequência da categoria de referência, que não é necessariamente o valor associado a essa categoria. Por exemplo, o seguinte subcomando CONTRAST obtém uma matriz de contraste que omite a segunda categoria:

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

Suponha que *fator* tenha quatro categorias. A matriz de contraste resultante será

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
```

```
( 0 -1 1 0)
( 0 -1 0 1)
```

## Helmert

**Contrastes de Helmert.** Compara as categorias de uma variável independente com a média das categorias subsequentes. O formulário matriz geral é

```
mean (1/k 1/k ... 1/k 1/k 1/k)
df(1) ( 1 -1/(k-1) ... -1/(k-1) -1/(k-1) -1/(k-1))
df(2) ( 0 1 ... -1/(k-2) -1/(k-2) -1/(k-2))
.
.
.
df(k-2) ( 0 0 ... 1 -1/2 -1/2)
df(k-1) ( 0 0 ... 0 1 -1)
```

em que  $k$  é o número de categorias da variável independente. Por exemplo, uma variável independente com quatro categorias possui uma matriz de contraste de Helmert da seguinte forma:

```
(1/4 1/4 1/4 1/4)
( 1 -1/3 -1/3 -1/3)
( 0 1 -1/2 -1/2)
( 0 0 1 -1)
```

## Diferença

**Diferença ou contrastes de Helmert reversos.** Compara as categorias de uma variável independente com a média das categorias anteriores da variável. O formulário matriz geral é

```
mean ( 1/k 1/k 1/k ... 1/k)
df(1) ( -1 1 0 ... 0)
df(2) ( -1/2 -1/2 1 ... 0)
.
.
.
df(k-1) (-1/(k-1) -1/(k-1) -1/(k-1) ... 1)
```

em que  $k$  é o número de categorias para a variável independente. Por exemplo, os contrastes de diferença para uma variável independente com quatro categorias são os seguintes:

```
( 1/4 1/4 1/4 1/4)
( -1 1 0 0)
(-1/2 -1/2 1 0)
(-1/3 -1/3 -1/3 1)
```

## Polinomial

**Contrastes polinomiais ortogonais.** O primeiro grau de liberdade contém o efeito linear em todas as categorias; o segundo grau de liberdade, o efeito quadrático; o terceiro grau de liberdade, o cúbico, e assim por diante, para o efeito de ordem mais alta.

É possível especificar o espaçamento entre os níveis de tratamento medido pela variável categórica fornecida. O espaçamento igual, que é o padrão, se você omitir a métrica, pode ser especificado como números inteiros consecutivos de 1 a  $k$ , em que  $k$  é o número de categorias. Se a variável *drug* possuir três categorias, o subcomando

```
/CONTRAST(DRUG)=POLYNOMIAL
```

é o mesmo que

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

O espaçamento igual não é sempre necessário, no entanto. Por exemplo, suponha que *drug* representa dosagens diferentes de um determinado medicamento para os três grupos. Se a dosagem administrada ao segundo grupo for duas vezes maior que a dosagem fornecida ao primeiro grupo, e a dosagem administrada ao terceiro grupo for três vezes maior que para o primeiro grupo, as categorias de tratamento serão igualmente espaçadas e uma métrica apropriada para esta situação consiste em números inteiros consecutivos:

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

Se, no entanto, a dosagem administrada para o segundo grupo for quatro vezes maior que a fornecida para o primeiro grupo, e a dosagem administrada para o terceiro grupo for sete vezes maior que a fornecida para o primeiro grupo, uma métrica apropriada seria

```
/CONTRAST(DRUG)=POLYNOMIAL(1,4,7)
```

Em ambos os casos, o resultado da especificação de contraste é que o primeiro grau de liberdade para *drug* contém o efeito linear dos níveis de dosagem e o segundo grau de liberdade contém o efeito quadrático.

Contrastes polinomiais são especialmente úteis em testes de tendências e para investigar a natureza das superfícies de resposta. Também é possível usar os contrastes polinomiais para executar o ajuste de curva não linear, como regressão curvilinear.

## Repetido

**Compara níveis adjacentes de uma variável independente.** O formulário matriz geral é

```
mean (1/k 1/k 1/k ... 1/k 1/k)
df(1) ( 1 -1 0 ... 0 0)
df(2) ( 0 1 -1 ... 0 0)
.
.
df(k-1) ( 0 0 0 ... 1 -1)
```

em que  $k$  é o número de categorias para a variável independente. Por exemplo, os contrastes repetidos para uma variável independente com quatro categorias são os seguintes:

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
( 0 1 -1 0)
( 0 0 1 -1)
```

Esses contrastes são úteis na análise de perfil e sempre que os escores de diferença são necessários.

## Especial

**Um contraste definido pelo usuário.** Permite a entrada de contrastes especiais na forma de matrizes quadradas com tantas linhas e colunas quanto existirem categorias da variável independente fornecida. Para MANOVA e LOGLINEAR, a primeira linha inserida sempre será a média ou constante, efeito e representa o conjunto de ponderações, indicando como obter a média de outras variáveis independentes, se houver, sobre a variável fornecida. Geralmente, esse contraste é um vetor de uns.

As linhas restantes da matriz contém os contrastes especial indicando as comparações entre as categorias da variável. Geralmente, contrastes ortogonais são os mais úteis. Os contrastes ortogonais são estatisticamente independentes e são não redundantes. Os contrastes são ortogonais se:

- Para cada linha, os coeficientes de contraste somarem 0.
- Os produtos de coeficientes correspondentes para todos os pares de linhas separadas também somarem 0.

Por exemplo, suponha que o tratamento possua quatro níveis e que você deseja comparar os vários níveis de tratamento uns com os outros. Um contraste especial apropriado é

```
(1 1 1 1) weights for mean calculation
(3 -1 -1 -1) compare 1st with 2nd through 4th
(0 2 -1 -1) compare 2nd with 3rd and 4th
(0 0 1 -1) compare 3rd with 4th
```

que é especificado por meio do subcomando CONTRAST a seguir para o MANOVA, LOGISTIC REGRESSION e o COXREG:

```
/CONTRAST(TREATMNT)=SPECIAL( 1 1 1 1
                             3 -1 -1 -1
                             0 2 -1 -1
                             0 0 1 -1 )
```

Para LOGLINEAR, é necessário especificar:

```
/CONTRAST(TREATMNT)=BASIS SPECIAL ( 1 1 1 1
3 -1 -1 -1
0 2 -1 -1
0 0 1 -1 )
```

Cada linha, exceto as somas de linha média para 0. Produtos de cada par de linhas separadas somam 0 também:

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 4: (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0
```

Os contrastes especiais não precisam ser ortogonais. No entanto, eles não devem ser combinações lineares uns dos outros. Se forem, o procedimento relata a dependência linear e interrompe o processamento. Helmert, diferença e contrastes polinomial são todos contrastes ortogonais.

## Indicador

**Codificação de variável indicadora.** Também conhecido como codificação simulada, esta codificação não está disponível em LOGLINEAR ou MANOVA. O número de novas variáveis codificadas é  $k-1$ . Casos na categoria de referência são codificados como 0 para todas as variáveis  $k-1$ . Um caso na categoria  $i^{\text{th}}$  é codificado como 0 para todas as variáveis indicadoras, exceto o  $i^{\text{th}}$ , que é codificado como 1.

## Estruturas de Covariâncias

Essa seção fornece informações adicionais sobre estruturas de covariâncias.

**Ante-Dependência: Primeira Ordem.** Essa estruturas de covariâncias tem variâncias heterogêneas e correlações heterogêneas entre elementos adjacentes. A correlação entre dois elementos não adjacentes é o produto das correlações entre os elementos que se encontram entre os elementos de interesse.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_1$	$\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_4\sigma_1\rho_1\rho_2\rho_3)$
$(\sigma_2\sigma_1\rho_1$	$\sigma_2^2$	$\sigma_3\sigma_2\rho_2$	$\sigma_4\sigma_2\rho_2\rho_3)$
$(\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_3\sigma_2\rho_2$	$\sigma_3^2$	$\sigma_4\sigma_3\rho_3)$
$(\sigma_4\sigma_1\rho_1\rho_2\rho_3$	$\sigma_4\sigma_2\rho_2\rho_3$	$\sigma_4\sigma_3\rho_3$	$\sigma_4^2)$

**AR(1).** Essa é a estrutura autorregressiva de primeira ordem com variâncias homogêneas. A correlação entre quaisquer dois elementos é igual a  $\rho$  para elementos adjacentes,  $\rho^2$  para elementos que são separados por um terceiro, e assim por diante. é restringido de forma que  $-1 < \rho < 1$ .

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2$	$\sigma^2\rho^3)$
$(\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2)$
$(\sigma^2\rho^2$	$\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho)$
$(\sigma^2\rho^3$	$\sigma^2\rho^2$	$\sigma^2\rho$	$\sigma^2)$

**AR(1): Heterogêneo.** Essa é a estrutura autorregressiva de primeira ordem com variâncias heterogêneas. A correlação entre quaisquer dois elementos é igual a  $\rho$  para elementos adjacentes,  $\rho^2$  para dois elementos separados por um terceiro, e assim por diante. é restringido para estar entre  $-1$  e  $1$ .

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho^2$	$\sigma_4\sigma_1\rho^3)$
$(\sigma_2\sigma_1\rho$	$\sigma_2^2$	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho^2)$
$(\sigma_3\sigma_1\rho^2$	$\sigma_3\sigma_2\rho$	$\sigma_3^2$	$\sigma_4\sigma_3\rho)$
$(\sigma_4\sigma_1\rho^3$	$\sigma_4\sigma_2\rho^2$	$\sigma_4\sigma_3\rho$	$\sigma_4^2)$

**ARMA(1,1).** Essa é a estrutura média móvel autorregressiva de primeira ordem. Ela possui variâncias homogêneas. A correlação entre dois elementos é igual a  $\rho$  para elementos adjacentes,  $\rho^2$  para elementos separados por um terço, e assim por diante.  $\rho$  e  $\sigma^2$  são os parâmetros autorregressivos e de média móvel, respectivamente, e os seus valores são restritos para estarem entre  $-1$  e  $1$ , inclusive.

$\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2$	$\sigma^2\rho^3$
$\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2$
$\sigma^2\rho^2$	$\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$
$\sigma^2\rho^3$	$\sigma^2\rho^2$	$\sigma^2\rho$	$\sigma^2$

**Simetria Composta.** Essa estrutura tem variância constante e covariância constante.

$(\sigma^2 + \sigma_1^2)$	$\sigma_1$	$\sigma_1$	$\sigma_1$
$\sigma_1$	$(\sigma^2 + \sigma_1^2)$	$\sigma_1$	$\sigma_1$
$\sigma_1$	$\sigma_1$	$(\sigma^2 + \sigma_1^2)$	$\sigma_1$
$\sigma_1$	$\sigma_1$	$\sigma_1$	$(\sigma^2 + \sigma_1^2)$

**Simetria Composta: Métrica de Correlação.** Essa estruturas de covariâncias tem variâncias homogêneas e correlações homogêneas entre elementos.

$\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$
$\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$
$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$
$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2$

**Simetria Composta: Heterogênea.** Essa estruturas de covariâncias tem variâncias heterogêneas e correlação constante entre elementos.

$(\sigma_1^2)$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho$	$\sigma_4\sigma_1\rho$
$\sigma_2\sigma_1\rho$	$(\sigma_2^2)$	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho$
$\sigma_3\sigma_1\rho$	$\sigma_3\sigma_2\rho$	$(\sigma_3^2)$	$\sigma_4\sigma_3\rho$
$\sigma_4\sigma_1\rho$	$\sigma_4\sigma_2\rho$	$\sigma_4\sigma_3\rho$	$(\sigma_4^2)$

**Diagonal.** Essa estruturas de covariâncias tem variâncias heterogêneas e zero correlação entre elementos.

$(\sigma_1^2)$	0	0	0
0	$(\sigma_2^2)$	0	0
0	0	$(\sigma_3^2)$	0
0	0	0	$(\sigma_4^2)$

**AR1 de produto direto (UN\_AR1).** Especifica o produto Kronecker de uma matriz não estruturada e a outra matriz de covariâncias de autorregressão de primeira ordem. A primeira matriz não estruturada modela a observação multivariada e a segunda estrutura de covariâncias de autorregressão de primeira ordem modela a covariância de dados através do tempo ou de outro fator.

**Produto direto não estruturado (UN\_UN).** Especifica o produto Kronecker de duas matrizes não estruturadas, com a primeira modelando a observação multivariada e a segunda modelando a covariância de dados através do tempo ou de outro fator.

**Simetria composta de produto direto (UN\_CS).** Especifica o produto Kronecker de uma matriz não estruturada e a outra matriz de covariâncias de simetria composta com variância e covariância

constantes. A primeira matriz não estruturada modela a observação multivariada e a segunda estrutura de covariâncias de simetria composta modela a covariância de dados através do tempo ou de outro fator.

**Análítica de Fator: Primeira Ordem.** Essa estruturas de covariâncias tem variâncias heterogêneas que são compostas de um termo que é heterogêneo ao longo de elementos e um termo que é homogêneo ao longo de elementos. A covariância entre quaisquer dois elementos é a raiz quadrada do produto de seus termos de variância heterogênea.

$$\begin{pmatrix} \lambda_1^2 + d & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2 + d & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + d & \lambda_4\lambda_3 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2 + d \end{pmatrix}$$

**Análítica de Fator: Primeira Ordem, Heterogêneo.** Essa estruturas de covariâncias tem variâncias heterogêneas que são compostas de dois termos que são heterogêneos ao longo de elementos. A covariância entre quaisquer dois elementos é a raiz quadrada do produto do primeiro de seus termos de variância heterogênea.

$$\begin{pmatrix} \lambda_1^2 + d_1 & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2 + d_2 & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + d_3 & \lambda_4\lambda_3 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2 + d_4 \end{pmatrix}$$

**Huynh-Feldt.** Essa é uma matriz "circular" na qual a covariância entre quaisquer dois elementos é igual à média de suas variâncias menos uma constante. Nem as variâncias e nem as covariâncias são constantes.

$$\begin{pmatrix} \sigma_1^2 & [\sigma_1^2 + \sigma_2^2]/2 - \lambda & [\sigma_1^2 + \sigma_3^2]/2 - \lambda & [\sigma_1^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_2^2]/2 - \lambda & \sigma_2^2 & [\sigma_2^2 + \sigma_3^2]/2 - \lambda & [\sigma_2^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_3^2]/2 - \lambda & [\sigma_2^2 + \sigma_3^2]/2 - \lambda & \sigma_3^2 & [\sigma_3^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_4^2]/2 - \lambda & [\sigma_2^2 + \sigma_4^2]/2 - \lambda & [\sigma_3^2 + \sigma_4^2]/2 - \lambda & \sigma_4^2 \end{pmatrix}$$

**Identidade Escalada.** Essa estrutura tem variância constante. É assumido que não há nenhuma correlação entre qualquer elemento.

$$\begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

**Espacial: Potência.** Essa estruturas de covariâncias tem variâncias homogêneas e correlações heterogêneas entre elementos.  $d_{ij}$  é a distância euclidiana estimada entre a  $i^{\text{th}}$  e a  $j^{\text{th}}$  medidas.

$$\begin{pmatrix} \sigma^2 & \sigma^2 \rho^{d_{12}} & \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{14}} \\ \sigma^2 \rho^{d_{12}} & \sigma^2 & \sigma^2 \rho^{d_{23}} & \sigma^2 \rho^{d_{24}} \\ \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{23}} & \sigma^2 & \sigma^2 \rho^{d_{34}} \\ \sigma^2 \rho^{d_{14}} & \sigma^2 \rho^{d_{24}} & \sigma^2 \rho^{d_{34}} & \sigma^2 \end{pmatrix}$$

**Espacial: Exponencial.** Essa estruturas de covariâncias tem variâncias homogêneas e correlações heterogêneas entre elementos.  $d_{ij}$  é a distância euclidiana estimada entre a  $i^{\text{th}}$  e a  $j^{\text{th}}$  medidas.

$$\begin{pmatrix} \sigma^2 & \sigma^2 \exp\{-d_{12}/\theta\} & \sigma^2 \exp\{-d_{13}/\theta\} & \sigma^2 \exp\{-d_{14}/\theta\} \\ \sigma^2 \exp\{-d_{12}/\theta\} & \sigma^2 & \sigma^2 \exp\{-d_{23}/\theta\} & \sigma^2 \exp\{-d_{24}/\theta\} \\ \sigma^2 \exp\{-d_{13}/\theta\} & \sigma^2 \exp\{-d_{23}/\theta\} & \sigma^2 & \sigma^2 \exp\{-d_{34}/\theta\} \\ \sigma^2 \exp\{-d_{14}/\theta\} & \sigma^2 \exp\{-d_{24}/\theta\} & \sigma^2 \exp\{-d_{34}/\theta\} & \sigma^2 \end{pmatrix}$$

$(\sigma^2 \exp\{-d_{12}/\theta\}$	$\sigma^2$	$\sigma^2 \exp\{-d_{23}/\theta\}$	$\sigma^2 \exp\{-d_{24}/\theta\}$
$(\sigma^2 \exp\{-d_{13}/\theta\}$	$\sigma^2 \exp\{-d_{23}/\theta\}$	$\sigma^2$	$\sigma^2 \exp\{-d_{34}/\theta\}$
$(\sigma^2 \exp\{-d_{14}/\theta\}$	$\sigma^2 \exp\{-d_{24}/\theta\}$	$\sigma^2 \exp\{-d_{34}/\theta\}$	$\sigma^2$

**Espacial: Gaussiano.** Essa estruturas de covariâncias tem variâncias homogêneas e correlações heterogêneas entre elementos.  $d_{ij}$  é a distância euclidiana estimada entre a  $i^{\text{th}}$  e a  $j^{\text{th}}$  medidas.

$(\sigma^2$	$\sigma^2 \exp\{-d_{12}/\rho^2\}$	$\sigma^2 \exp\{-d_{13}/\rho^2\}$	$\sigma^2 \exp\{-d_{14}/\rho^2\}$
$(\sigma^2 \exp\{-d_{12}/\rho^2\}$	$\sigma^2$	$\sigma^2 \exp\{-d_{23}/\rho^2\}$	$\sigma^2 \exp\{-d_{24}/\rho^2\}$
$(\sigma^2 \exp\{-d_{13}/\rho^2\}$	$\sigma^2 \exp\{-d_{23}/\rho^2\}$	$\sigma^2$	$\sigma^2 \exp\{-d_{34}/\rho^2\}$
$(\sigma^2 \exp\{-d_{14}/\rho^2\}$	$\sigma^2 \exp\{-d_{24}/\rho^2\}$	$\sigma^2 \exp\{-d_{34}/\rho^2\}$	$\sigma^2$

**Espacial: Linear.** Essa estruturas de covariâncias tem variâncias homogêneas e correlações heterogêneas entre elementos.  $d_{ij}$  é a distância euclidiana estimada entre a  $i^{\text{th}}$  e  $j^{\text{th}}$  medidas, e  $1_{ij}$  é uma função indicadora que será 1 se  $pd_{ij} \leq 0$  e for diferente de 0.

$(\sigma^2$	$\sigma^2(1 - pd_{12}) 1_{12}$	$\sigma^2(1 - pd_{13}) 1_{13}$	$\sigma^2(1 - pd_{14}) 1_{14}$
$(\sigma^2(1 - pd_{12}) 1_{12}$	$\sigma^2$	$\sigma^2(1 - pd_{23}) 1_{23}$	$\sigma^2(1 - pd_{24}) 1_{24}$
$(\sigma^2(1 - pd_{13}) 1_{13}$	$\sigma^2(1 - pd_{23}) 1_{23}$	$\sigma^2$	$\sigma^2(1 - pd_{34}) 1_{34}$
$(\sigma^2(1 - pd_{14}) 1_{14}$	$\sigma^2(1 - pd_{24}) 1_{24}$	$\sigma^2(1 - pd_{34}) 1_{34}$	$\sigma^2$

**Espacial: Log linear.** Essa estruturas de covariâncias tem variâncias homogêneas e correlações heterogêneas entre elementos.  $d_{ij}$  é a distância euclidiana estimada entre a  $i^{\text{th}}$  e  $j^{\text{th}}$  medidas e  $1_{ij}$  é uma função indicadora que será 1 se  $p \log(d_{ij}) \leq 0$  e for diferente de 0.

$(\sigma^2$	$\sigma^2(1 - p \log(d_{12})) 1_{12}$	$\sigma^2(1 - p \log(d_{13})) 1_{13}$	$\sigma^2(1 - p \log(d_{14})) 1_{14}$
$(\sigma^2(1 - p \log(d_{12})) 1_{12}$	$\sigma^2$	$\sigma^2(1 - p \log(d_{23})) 1_{23}$	$\sigma^2(1 - p \log(d_{24})) 1_{24}$
$(\sigma^2(1 - p \log(d_{13})) 1_{13}$	$\sigma^2(1 - p \log(d_{23})) 1_{23}$	$\sigma^2$	$\sigma^2(1 - p \log(d_{34})) 1_{34}$
$(\sigma^2(1 - p \log(d_{14})) 1_{14}$	$\sigma^2(1 - p \log(d_{24})) 1_{24}$	$\sigma^2(1 - p \log(d_{34})) 1_{34}$	$\sigma^2$

**Espacial: Esférico.** Essa estruturas de covariâncias tem variâncias homogêneas e correlações heterogêneas entre elementos.  $r_{ij} = d_{ij}/\rho$ , em que  $d_{ij}$  é a distância euclidiana estimada entre a  $i^{\text{th}}$  e  $j^{\text{th}}$  medidas.  $1_{ij}$  é uma função indicadora que será 1 se  $d_{ij} \leq \rho$  e for diferente de 0.

$(\sigma^2$	$\sigma^2(1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12}$	$\sigma^2(1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13}$	$\sigma^2(1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14}$
$(\sigma^2(1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12}$	$\sigma^2$	$\sigma^2(1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23}$	$\sigma^2(1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24}$
$(\sigma^2(1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13}$	$\sigma^2(1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23}$	$\sigma^2$	$\sigma^2(1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34}$
$(\sigma^2(1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14}$	$\sigma^2(1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24}$	$\sigma^2(1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34}$	$\sigma^2$

**Toeplitz.** Essa estruturas de covariâncias tem variâncias homogêneas e correlações heterogêneas entre elementos. A correlação entre elementos adjacentes é homogênea em pares de elementos adjacentes. A correlação entre elementos separados por um terceiro é novamente homogênea, e assim por diante.

$(\sigma^2$	$\sigma^2\rho_1$	$\sigma^2\rho_2$	$\sigma^2\rho_3$
$(\sigma^2\rho_1$	$\sigma^2$	$\sigma^2\rho_1$	$\sigma^2\rho_2$
$(\sigma^2\rho_2$	$\sigma^2\rho_1$	$\sigma^2$	$\sigma^2\rho_1$

$(\sigma^2\rho_3$	$\sigma^2\rho_2$	$\sigma^2\rho_1$	$\sigma^2)$
-------------------	------------------	------------------	-------------

**Toeplitz: Heterogênea.** Essa estruturas de covariâncias tem variâncias heterogêneas e correlações heterogêneas entre elementos. A correlação entre elementos adjacentes é homogênea em pares de elementos adjacentes. A correlação entre elementos separados por um terceiro é novamente homogênea, e assim por diante.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_1$	$\sigma_3\sigma_1\rho_2$	$\sigma_4\sigma_1\rho_3)$
$(\sigma_2\sigma_1\rho_1$	$\sigma_2^2$	$\sigma_3\sigma_2\rho_1$	$\sigma_4\sigma_2\rho_2)$
$(\sigma_3\sigma_1\rho_2$	$\sigma_3\sigma_2\rho_1$	$\sigma_3^2$	$\sigma_4\sigma_3\rho_1)$
$(\sigma_4\sigma_1\rho_3$	$\sigma_4\sigma_2\rho_2$	$\sigma_4\sigma_3\rho_1$	$\sigma_4^2)$

**Não estruturada.** Essa é um matriz de covariâncias completamente geral.

$(\sigma_1^2$	$2\ 1$	$3\ 1$	$4\ 1)$
$(\ 2\ 1$	$\sigma_2^2$	$\sigma_{32}$	$\sigma_{4\ 2})$
$(\sigma_{31}$	$\sigma_{32}$	$\sigma_3^2$	$\sigma_{4\ 3})$
$(\sigma_{41}$	$\sigma_{4\ 2}$	$\sigma_{4\ 3}$	$\sigma_4^2)$

**Não estruturada: Métrica de Correlação.** Essa estruturas de covariâncias tem variâncias heterogêneas e correlações heterogêneas.

$(\sigma_1^2$	$2\ 1\ 21$	$3\ 1\ 31$	$4\ 1\ 41)$
$(\ 2\ 1\ 21$	$\sigma_2^2$	$3\ 2\ 32$	$4\ 2\ 42)$
$(\ 3\ 1\ 31$	$3\ 2\ 32$	$\sigma_3^2$	$4\ 3\ 43)$
$(\ 4\ 1\ 41$	$4\ 2\ 42$	$4\ 3\ 43$	$\sigma_4^2)$

**Componentes de Variância.** Essa estrutura atribui uma estrutura de identidade escalada (ID) a cada um dos efeitos aleatórios especificados.

## Estatísticas bayesianas

IBM SPSS Estatísticas fornece suporte para as estatísticas bayesianas a seguir.

### Testes T de uma amostra e de amostra pareada

O procedimento Inferência bayesiana de uma amostra fornece opções para a execução da inferência bayesiana em um teste t de par de uma amostra e de duas amostras caracterizando distribuições posteriores. Quando os dados são normais, é possível usar um anterior normal para obter um posterior normal.

### Testes de proporção binomial

A inferência bayesiana de uma amostra: o procedimento binômio que fornece opções para a execução da inferência bayesiana de uma amostra na distribuição binomial. O parâmetro de interesse é  $\pi$ , que denota a probabilidade de êxito de um número fixo de avaliações que podem levar ao êxito ou à falha. Observe que as avaliações são independentes entre si e a probabilidade permanece a mesma em cada avaliação. Uma variável aleatória binômica pode ser vista como a soma de um número fixo de avaliações de Bernoulli independentes.

### Análise de distribuição de Poisson

Inferência bayesiana de uma amostra: O procedimento de Poisson fornece opções para executar a inferência bayesiana de uma amostra na distribuição de Poisson. A distribuição de Poisson, que é um modelo útil para eventos raros, supõem que, dentro de intervalos de tempo curtos, a probabilidade de ocorrência de um evento é proporcional ao período de tempo de espera. Um conjugado anterior da família de distribuição Gama é usado ao desenhar a inferência estatística bayesiana na distribuição de Poisson.



### **Amostras relacionadas**

O design de inferência de amostra bayesiano relacionado é bastante semelhante à inferência bayesiana de uma amostra em termos de tratamento de amostras emparelhadas. É possível especificar os nomes de variáveis em pares e executar a análise bayesiana na diferença média.

### **Testes T de amostras independentes**

O procedimento de inferência bayesiana de amostra independente fornece opções para o uso de uma variável de grupo para a definição de dois grupos independentes e a execução da inferência bayesiana na diferença das médias dos dois grupos. É possível estimar os fatores bayesianos utilizando abordagens diferentes e também caracterizar a distribuição posterior desejada assumindo que as variações são conhecidas ou desconhecidas.

### **Correlação entre pares (Pearson)**

A inferência bayesiana sobre o coeficiente de correlação de Pearson mede a relação linear entre duas variáveis de escala que seguem juntas uma distribuição normal bivariada. A inferência estatísticas convencional sobre o coeficiente de correlação foi amplamente discutida e sua prática é oferecida há muito tempo no IBM SPSS Statistics. O design da inferência bayesiana sobre o coeficiente de correlação de Pearson permite desenhar a inferência bayesiana estimando fatores de Bayes e caracterizando as distribuições posteriores.

### **Regressão linear**

A inferência bayesiana sobre regressão linear é um método estatístico amplamente utilizado na modelagem quantitativa. A regressão linear é uma abordagem básica e padrão, na qual os pesquisadores usam os valores de diversas variáveis para explicar ou prever valores de um resultado de escala. A regressão linear bayesiana univariada é uma abordagem da regressão linear na qual a análise estatística é executada no contexto da inferência bayesiana.

### **ANOVA de um fator**

O procedimento bayesiano ANOVA de um fator produz uma análise de variância para um fator de uma variável dependente quantitativa por uma variável de fator único (independente). A análise de variância é utilizada para testar a hipótese de que várias médias são iguais. O SPSS Statistics suporta fatores de Bayes, anteriores conjugados e anteriores não informativos.

### **Modelos de regressão log-linear**

O projeto para testar a independência de dois fatores requer duas variáveis categóricas para a construção de uma tabela de contingência e faz a inferência bayesiana na associação entre linha e coluna. É possível estimar os fatores bayesianos assumindo diferentes modelos e caracterizar a distribuição posterior desejada simulando o intervalo crível simultâneo para os termos de interação.

### **ANOVA de medidas repetidas de um fator**

O procedimento ANOVA de medidas repetidas de um fator bayesiano mede um fator do mesmo assunto a cada momento ou condição distinta e permite que os assuntos sejam cruzados dentro dos níveis. Supõe-se que cada assunto tenha uma única observação para cada momento ou condição (portanto, a interação de assunto-tratamento não é considerada).

## **Inferência de uma amostra bayesiana: normal**

Este recurso requer a opção SPSS Statistics Standard Edition ou Advanced Statistics.

O procedimento Inferência bayesiana de uma amostra: normal fornece opções para a execução da inferência bayesiana em testes t com pares de uma amostra e de duas amostras, caracterizando distribuições posteriores. Quando os dados são normais, é possível usar um anterior normal para obter um posterior normal.

1. Nos menus, escolha:

**Analisar > Estatísticas bayesianas > Uma amostra normal**

2. Selecione as **Variáveis de teste** apropriadas da lista **Variáveis disponíveis**. Pelo menos uma variável deve ser selecionada.

**Nota:** A lista de variáveis disponíveis fornece todas as variáveis, exceto as variáveis de Data e Sequência.

3. Selecione a **Análise bayesiana:**

- **Caracterizar distribuição posterior:** quando selecionado, a inferência bayesiana é feita de uma perspectiva que é abordada caracterizando as distribuições posteriores. É possível investigar a distribuição marginal posterior dos parâmetros de interesse, integrando os outros parâmetros inconvenientes e construir melhor os intervalos de credibilidade para fazer uma inferência direta. Esta é a definição padrão.
- **Estimar fator Bayes:** quando selecionado, a estimativa de fatores Bayes (uma das metodologias notáveis na inferência bayesiana) constitui uma razão natural para comparar as probabilidades marginais entre um nulo e uma hipótese alternativa.

*Tabela 2. Limites comumente usados para definir a significância de evidência*

Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
>100	Evidência extrema para H0	1-3	Evidência anedótica para H0	1/30-1/10	Evidência forte para H1
30-100	Evidência muito forte para H0	1	Sem evidência	1/100-1/30	Evidência muito forte para H1
10-30	Evidência forte para H0	1/3-1	Evidência anedótica para H1	1/100	Evidência extrema para H1
3-10	Evidência moderada para H0	1/10-1/3	Evidência moderada para H1		

*H0: Hipótese nula*

*H1: Hipótese alternativa*

1

2

- **Usar ambos os métodos:** quando selecionado, ambos os métodos de inferência, **Caracterizar distribuição posterior** e **Estimar fator Bayes**, são usados.
4. Selecione e/ou insira as configurações **Variação de dados e Valores de hipótese** apropriadas. A tabela reflete as variáveis que estão atualmente na lista **Variáveis de teste**. À medida em que as variáveis são incluídas ou removidas da lista **Variáveis de teste**, a tabela automaticamente inclui ou remove as mesmas variáveis de suas colunas de variáveis.
- Quando uma ou mais variáveis estão na lista **Variáveis de teste**, as colunas **Variável conhecida** e **Valor de variação** são ativadas.
- Variância conhecida**  
Selecione esta opção para cada variável quando a variação for conhecida.
- Valor da variância**  
Um parâmetro opcional que especifica o valor de variação, se conhecido, para os dados observados.
- Quando uma ou mais variáveis estão na lista **Variáveis de teste** e **Caracterizar distribuição posterior** não está selecionada, as colunas **Valor de teste nulo** e **Valor g** são ativadas.

<sup>1</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>2</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### Valor de teste nulo

Um parâmetro necessário que especifica o valor nulo na estimação do fator de Bayes. Somente um valor é permitido e 0 é o valor padrão.

### Valor de $g$

Especifica o valor a ser definido  $\psi^2 = g\sigma^2_x$  na estimativa do fator bayesiano. Quando o **Valor de variância** é especificado, o **Valor  $g$**  é padronizado para 1. Quando o **Valor de variância** não for especificado, é possível especificar um  $g$  fixo ou omitir o valor para integrá-lo.

5. Opcionalmente, é possível clicar em **Critérios** para especificar as configurações de “Inferência de uma amostra bayesiana: critérios” na página 109 (porcentagem de intervalo crível, opções de valores omissos e configurações de método numérico) ou clicar em **Anteriores** para especificar as configurações de “Inferência de uma amostra bayesiana: anteriores normais” na página 110 (tipo de anteriores, como parâmetros de inferência, variância média especificada ou precisão).

## Inferência de uma amostra bayesiana: critérios

É possível especificar os seguintes critérios de análise para sua Inferência de uma amostra bayesiana:

### Porcentagem % de intervalo crível

Especifique o nível de significância para os intervalos críveis de cálculo. O nível padrão é 95%.

### Valores omissos

Especifique o método no qual controlar valores omissos.

#### Excluir casos por método pairwise

Esta é a configuração padrão e exclui registros com valores omissos em uma base análise por análise. Os registros que incluem valores omissos, para um campo que é usado por um teste específico, são omitidos do teste.

#### Excluir casos pelo método listwise

Essa configuração exclui os registros que incluem listwise de valores omissos. Os registros que incluem valores omissos para qualquer campo que é nomeado em qualquer subcomando são excluídos de todas as análises.

**Nota:** As opções a seguir estão disponíveis somente quando a opção **Estimar fator bayesiano** ou **Usar ambos os métodos** está selecionada para **Análise bayesiana**.

### Método Numérico

Especifique o método numérico que é usado para estimar o integral.

#### Quadratura Gauss-Lobatto Adaptável

Esta é a configuração padrão e chama a abordagem Quadratura Gauss-Lobatto Adaptável.

#### Tolerância

Especifique o valor de tolerância para os métodos numéricos. A configuração padrão é 0.000001. A opção está disponível apenas quando a configuração **Quadratura Gauss-Lobatto Adaptável** é selecionada.

#### Máximo de iterações

Especifique o número máximo de iterações do método Quadratura Gauss-Lobatto Adaptável. O valor deve ser um número inteiro positivo. A configuração padrão é 2000. A opção está disponível apenas quando a configuração **Quadratura Gauss-Lobatto Adaptável** é selecionada.

### Aproximação de Monte Carlo

Esta opção chama a abordagem Aproximação de Monte Carlo.

#### Configurar valor inicial customizado

Quando selecionada, é possível especificar um valor inicial customizado no campo **Valor inicial**.

#### Semente

Especifique um valor inicial aleatório configurado para o método de Aproximação de Monte Carlo. O valor deve ser um número inteiro positivo. Por padrão, um valor inicial aleatório é designado.

### Número de amostras Monte Carlo

Especifique o número de pontos que servem de amostra para a aproximação de Monte Carlo. O valor deve ser um número inteiro positivo. O valor padrão é 1000000. A opção fica disponível somente quando a configuração **Aproximação de Monte Carlo** é selecionada.

## Inferência de uma amostra bayesiana: anteriores normais

É possível especificar os seguintes critérios de distribuição anteriores para sua Inferência de uma amostra bayesiana:

**Nota:** Muitos pesquisadores aplicados podem questionar a necessidade de especificar a priori. Os a priori de referência minimizam a preocupação de onde o a priori geralmente fica sobrecarregado conforme os dados aumentam. Quando informações de a priori informativo são especificadas, os métodos bayesianos podem efetivamente usar as informações. A necessidade de especificar um a priori não deve ser vista como um impedimento para usar a análise bayesiana.

### Anterior na variância/precisão

Fornece opções para a definição de valores de variância e de precisão.

#### Variância

Selecione para especificar a distribuição anterior para o parâmetro de variância. Quando essa opção é selecionada, a lista **Distribuição anterior** fornece as seguintes opções:

**Nota:** Quando a variância de dados já está especificado para algumas variáveis, as configurações a seguir são ignoradas para essas variáveis.

- **Difuso** - a configuração padrão. Especifica os parâmetros a priori difusos.
- **Qui-quadrado inverso** - Especifica a distribuição e parâmetros para  $\chi^2(v_0, \sigma^2_0)$  inverso, em que  $v_0 > 0$  é o grau de liberdade e  $\sigma^2_0 > 0$  é o parâmetro de escala.
- **Gama inversa** - Especifica a distribuição e parâmetros para Gama inversa( $\alpha_0, \beta_0$ ), em que  $\alpha_0 > 0$  é o parâmetro de forma e  $\beta_0 > 0$  é o parâmetro de escala.
- **Jeffreys S2** - Especifica o a priori não informativo  $\propto 1/\sigma^2_0$ .
- **Jeffreys S4** - Especifica o a priori não informativo  $\propto 1/\sigma^4_0$ .

#### Precisão

Selecione para especificar a distribuição anterior para o parâmetro de precisão. Quando essa opção é selecionada, a lista **Distribuição anterior** fornece as seguintes opções:

- **Gama** - Especifica a distribuição e parâmetros para Gama ( $\alpha_0, \beta_0$ ), em que  $\alpha_0 > 0$  é o parâmetro de forma e  $\beta_0 > 0$  é o parâmetro de escala.
- **Qui-quadrado** - Especifica a distribuição e parâmetros para  $\chi^2(v_0)$ , em que  $v_0 > 0$  é o grau de liberdade.

#### Parâmetro de Forma

Especifique o parâmetro de forma  $a_0$  para a distribuição Inverse-Gamma. Deve-se inserir um valor único maior que 0.

#### Parâmetro de escala

Especifique o parâmetro de escala  $b_0$  para a distribuição Inverse-Gamma. Deve-se inserir um valor único maior que 0. Quanto maior o parâmetro de escala, mais difusa a distribuição.

### Anterior na média fornecida de variância/precisão

Especifique a distribuição anterior para o parâmetro de média que é condicional em relação à variância ou ao parâmetro de precisão.

#### Normal

Especifica a distribuição e parâmetros para Normal( $\mu_0, K^{-1}_0\sigma^2_0$ ) na variância ou Normal( $\mu_0, K_0/\sigma^2_0$ ) na precisão, em que  $\mu_0 \in (-\infty, \infty)$  e  $\sigma^2 > 0$ .

### Parâmetro de Localização

Insira um valor numérico que especifique o parâmetro de local para a distribuição.

### Parâmetro de escala

Especifique o parâmetro de escala  $b_0$  para a distribuição Inverse-Gamma. Deve-se inserir um valor único maior que 0.

### Kappa

Especifique o valor de  $K_0$  em  $\text{Normal}(\mu_0, K_0^{-1}\sigma_0^2)$  ou  $\text{Normal}(\mu_0, K_0/\sigma_0^2)$ . Deve-se inserir um valor único maior que 0 (1 é o valor padrão).

### Difuso

A configuração padrão que especifica o anterior difuso  $\propto 1$ .

## Inferência de uma amostra bayesiana: binômio

Este recurso requer a opção SPSS Statistics Standard Edition ou Advanced Statistics.

A inferência bayesiana de uma amostra: o procedimento binômio que fornece opções para a execução da inferência bayesiana de uma amostra na distribuição binomial. O parâmetro de interesse é  $\pi$ , que denota a probabilidade de êxito de um número fixo de avaliações que podem levar ao êxito ou à falha. Observe que as avaliações são independentes entre si e a probabilidade permanece a mesma em cada avaliação. Uma variável aleatória binômica pode ser vista como a soma de um número fixo de avaliações de Bernoulli independentes.

Embora isso não seja necessário, normalmente um anterior da família de distribuição beta é escolhido ao estimar um parâmetro binômio. A família beta é conjugada para a família binomial e, como tal, leva à distribuição posterior com uma forma fechada ainda na família de distribuição beta.

1. Nos menus, escolha:

**Analisar > Estatísticas bayesianas > Binomial de uma amostra**

2. Selecione as **Variáveis de teste** apropriadas da lista **Variáveis disponíveis**. Pelo menos uma variável deve ser selecionada.

**Nota:** A lista de variáveis disponíveis fornece todas as variáveis, exceto as variáveis de Data e Sequência.

3. Selecione a **Análise bayesiana:**

- **Caracterizar distribuição posterior:** quando selecionado, a inferência bayesiana é feita de uma perspectiva que é abordada caracterizando as distribuições posteriores. É possível investigar a distribuição marginal posterior dos parâmetros de interesse, integrando os outros parâmetros inconvenientes e construir melhor os intervalos de credibilidade para fazer uma inferência direta. Esta é a definição padrão.
- **Estimar fator Bayes:** quando selecionado, a estimativa de fatores Bayes (uma das metodologias notáveis na inferência bayesiana) constitui uma razão natural para comparar as probabilidades marginais entre um nulo e uma hipótese alternativa.

*Tabela 3. Limites comumente usados para definir a significância de evidência*

Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
>100	Evidência extrema para H0	1-3	Evidência anedótica para H0	1/30-1/10	Evidência forte para H1
30-100	Evidência muito forte para H0	1	Sem evidência	1/100-1/30	Evidência muito forte para H1

Tabela 3. Limites comumente usados para definir a significância de evidência (continuação)					
Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
10-30	Evidência forte para H0	1/3-1	Evidência anedótica para H1	1/100	Evidência extrema para H1
3-10	Evidência moderada para H0	1/10-1/3	Evidência moderada para H1		

*H0: Hipótese nula*

*H1: Hipótese alternativa*

3

4

- **Usar ambos os métodos:** quando selecionado, ambos os métodos de inferência, **Caracterizar distribuição posterior** e **Estimar fator Bayes**, são usados.
4. Selecione e/ou insira as configurações adequadas para **Categorias de êxito e valores de hipótese**. A tabela reflete as variáveis que estão atualmente na lista **Variáveis de teste**. Quando as variáveis são incluídas ou removidas da lista **Variáveis de teste**, a tabela automaticamente inclui ou remove as mesmas variáveis de suas colunas pareadas de variáveis.
- Quando **Caracterizar distribuição posterior** é selecionado como a **Análise bayesiana**, a coluna **Categorias de êxito** é ativada.
  - Quando as opções **Estimar fator bayesiano** ou **Usar ambos os métodos** são selecionadas como **Análise bayesiana**, todas as colunas editáveis são ativadas.

#### Ponto nulo

Ativa e desativa a opção **Proporção nula**. Quando a configuração é ativada, as opções **Forma de a priori nulo** e **Escala de a priori nulo** são desativadas.

#### Forma anterior nula

Especifica o parâmetro de forma  $\alpha_0$  sob a hipótese nula de inferência binomial.

#### Escala anterior nula

Especifica o parâmetro de escala  $b_0$  sob a hipótese nula de inferência binomial.

#### Proporção nula

Especifica o parâmetro de forma  $\alpha_0$  e o parâmetro de escala  $b_0$  na hipótese nula para uma distribuição de a priori conjugada (para acomodar os parâmetros a priori Beta e de Haldane). O intervalo válido são valores numéricos entre 0 e 1.

#### Forma anterior alternativa

Um parâmetro necessário para especificar  $\alpha_0$  sob a hipótese alternativa de inferência binomial caso o fator de Bayes deva ser estimado.

#### Escala anterior alternativa

Um parâmetro necessário para especificar  $b_0$  sob a hipótese alternativa de inferência binomial caso o fator de Bayes deva ser estimado.

#### Categorias de êxito

Fornece opções para a definição de distribuições anteriores conjugadas. As opções fornecidas especificam como o êxito é definido, para variáveis numéricas e de sequência, quando os valores dos dados são testados em relação ao valor de teste.

<sup>3</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>4</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

**Última categoria**

A configuração padrão que executa o teste binomial utilizando o último valor numérico encontrado na categoria depois de classificada em ordem crescente.

**Primeira categoria**

Executa o teste binomial utilizando o primeiro valor numérico encontrado na categoria depois de classificada em ordem crescente.

**Ponto médio**

Usa como casos os valores numéricos  $\geq$  ao ponto médio. Um valor de ponto médio é a média dos dados de amostra mínimo e máximo.

**Ponto de corte**

Usa os valores numéricos  $\geq$  um valor de corte especificado como casos. A configuração deve ser um valor numérico único.

**Nível**

Trata como casos os valores de sequência especificados pelo usuário (é possível usar mais de um). Use vírgulas para separar os diferentes valores.

5. Opcionalmente, é possível clicar em **Crítérios** para especificar configurações de “Inferência de uma amostra bayesiana: critérios” na página 109 (porcentagem de intervalo crível, opções de valores ausentes e configurações de métodos numéricos) ou clicar em **Anteriores** para especificar configurações de “Inferência de uma amostra bayesiana: anteriores binomiais/de Poisson” na página 113 (distribuições anteriores conjugadas ou customizadas).

## Inferência de uma amostra bayesiana: anteriores binomiais/de Poisson

É possível especificar os seguintes critérios de distribuição anteriores para sua Inferência de uma amostra bayesiana:

**Nota:** Muitos pesquisadores aplicados podem questionar a necessidade de especificar a priori. Os a priori de referência minimizam a preocupação de onde o a priori geralmente fica sobrecarregado conforme os dados aumentam. Quando informações de a priori informativo são especificadas, os métodos bayesianos podem efetivamente usar as informações. A necessidade de especificar um a priori não deve ser vista como um impedimento para usar a análise bayesiana.

**Parâmetro de Forma**

Para parâmetros a priori binomiais, especifique o parâmetro de forma  $\alpha_0$  para distribuição Beta.

Para parâmetros a priori de Poisson, especifique o parâmetro de forma  $\alpha_0$  para distribuição Gama.

Deve-se inserir um valor único maior que 0.

**Parâmetro de escala**

Para parâmetros a priori binomiais, especifique o parâmetro de escala  $b_0$  para distribuição Beta.

Para parâmetros a priori de Poisson, especifique o parâmetro de escala  $b_0$  para distribuição Gama.

Deve-se inserir um valor único maior que 0.

## Inferência de uma amostra bayesiana: Poisson

Este recurso requer a opção SPSS Statistics Standard Edition ou Advanced Statistics.

Inferência bayesiana de uma amostra: O procedimento de Poisson fornece opções para executar a inferência bayesiana de uma amostra na distribuição de Poisson. A distribuição de Poisson, que é um modelo útil para eventos raros, supõem que, dentro de intervalos de tempo curtos, a probabilidade de ocorrência de um evento é proporcional ao período de tempo de espera. Um conjugado anterior da família de distribuição Gama é usado ao desenhar a inferência estatística bayesiana na distribuição de Poisson.

1. Nos menus, escolha:

**Analisar > Estatísticas bayesianas > Poisson de uma amostra**

2. Selecione as **Variáveis de teste** apropriadas da lista **Variáveis disponíveis**. Pelo menos uma variável deve ser selecionada.

**Nota:** A lista de variáveis disponíveis fornece todas as variáveis, exceto as variáveis de Data e Sequência.

3. Selecione a **Análise bayesiana**:

- **Caracterizar distribuição posterior:** quando selecionado, a inferência bayesiana é feita de uma perspectiva que é abordada caracterizando as distribuições posteriores. É possível investigar a distribuição marginal posterior dos parâmetros de interesse, integrando os outros parâmetros inconvenientes e construir melhor os intervalos de credibilidade para fazer uma inferência direta. Esta é a definição padrão.
- **Estimar fator Bayes:** quando selecionado, a estimativa de fatores Bayes (uma das metodologias notáveis na inferência bayesiana) constitui uma razão natural para comparar as probabilidades marginais entre um nulo e uma hipótese alternativa.

Tabela 4. Limites comumente usados para definir a significância de evidência					
Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
>100	Evidência extrema para H0	1-3	Evidência anedótica para H0	1/30-1/10	Evidência forte para H1
30-100	Evidência muito forte para H0	1	Sem evidência	1/100-1/30	Evidência muito forte para H1
10-30	Evidência forte para H0	1/3-1	Evidência anedótica para H1	1/100	Evidência extrema para H1
3-10	Evidência moderada para H0	1/10-1/3	Evidência moderada para H1		

*H0: Hipótese nula*

*H1: Hipótese alternativa*

5

6

- **Usar ambos os métodos:** quando selecionado, ambos os métodos de inferência, **Caracterizar distribuição posterior** e **Estimar fator Bayes**, são usados.

4. Selecione e/ou insira as configurações adequadas de **Valores de hipótese**. A tabela reflete as variáveis que estão atualmente na lista **Variáveis de teste**. Quando as variáveis são incluídas ou removidas da lista **Variáveis de teste**, a tabela automaticamente inclui ou remove as mesmas variáveis de suas colunas pareadas de variáveis.

- Quando **Caracterizar distribuição posterior** é selecionado como a **Análise bayesiana**, nenhuma das colunas é ativada.
- Quando as opções **Estimar fator bayesiano** ou **Usar ambos os métodos** são selecionadas como **Análise bayesiana**, todas as colunas editáveis são ativadas.

**Ponto nulo**

Ativa e desativa a opção **Taxa nula**. Quando a configuração é ativada, as opções **Forma de a priori nulo** e **Escala de a priori nulo** são desativadas.

<sup>5</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>6</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.



**Forma anterior nula**

Especifica o parâmetro de forma  $\alpha_0$  sob a hipótese nula da inferência de Poisson.

**Escala anterior nula**

Especifica o parâmetro de escala  $b_0$  sob a hipótese nula da inferência de Poisson.

**Taxa nula**

Especifica o parâmetro de forma  $\alpha_0$  e o parâmetro de escala  $b_0$  na hipótese nula para uma distribuição de a priori conjugado (para acomodar o relacionamento Poisson-Gama). O valor mínimo deve ser um valor numérico maior que 0; o valor máximo deve ser um valor duplo máximo.

**Forma anterior alternativa**

Um parâmetro necessário para especificar  $\alpha_1$  sob a hipótese alternativa da inferência de Poisson caso o fator de Bayes deva ser estimado.

**Escala anterior alternativa**

Um parâmetro necessário para especificar  $b_1$  sob a hipótese alternativa da inferência de Poisson caso o fator de Bayes deva ser estimado.

5. Opcionalmente, é possível clicar em **Crítérios** para especificar configurações de “Inferência de uma amostra bayesiana: critérios” na página 109 (porcentagem de intervalo crível, opções de valores ausentes e configurações de métodos numéricos) ou clicar em **Anteriores** para especificar configurações de “Inferência de uma amostra bayesiana: anteriores binomiais/de Poisson” na página 113 (distribuições anteriores conjugadas ou customizadas).

## Inferência de amostra relacionada bayesiana: normal

Este recurso requer a opção SPSS Statistics Standard Edition ou Advanced Statistics.

O procedimento Inferência bayesiana de amostra relacionada: normal fornece opções de inferência bayesiana de uma amostra para amostras com pares. É possível especificar os nomes de variáveis em pares e executar a análise bayesiana na diferença média.

1. Nos menus, escolha:

**Analisar > Estatísticas bayesianas > Amostras relacionadas normal**

2. Selecione as **Variáveis pareadas** apropriadas da lista **Variáveis disponíveis**. Pelo menos um par de variáveis de origem deve ser selecionado e não é possível selecionar mais de duas variáveis de origem para qualquer par configurado.

**Nota:** A lista de variáveis disponíveis fornece todas as variáveis, exceto as variáveis de Sequência.

3. Selecione a **Análise bayesiana:**

- **Caracterizar distribuição posterior:** quando selecionado, a inferência bayesiana é feita de uma perspectiva que é abordada caracterizando as distribuições posteriores. É possível investigar a distribuição marginal posterior dos parâmetros de interesse, integrando os outros parâmetros inconvenientes e construir melhor os intervalos de credibilidade para fazer uma inferência direta. Esta é a definição padrão.
- **Estimar fator Bayes:** quando selecionado, a estimativa de fatores Bayes (uma das metodologias notáveis na inferência bayesiana) constitui uma razão natural para comparar as probabilidades marginais entre um nulo e uma hipótese alternativa.

Tabela 5. Limites comumente usados para definir a significância de evidência					
Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
>100	Evidência extrema para H0	1-3	Evidência anedótica para H0	1/30-1/10	Evidência forte para H1
30-100	Evidência muito forte para H0	1	Sem evidência	1/100-1/30	Evidência muito forte para H1

Tabela 5. Limites comumente usados para definir a significância de evidência (continuação)					
Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
10-30	Evidência forte para H0	1/3-1	Evidência anedótica para H1	1/100	Evidência extrema para H1
3-10	Evidência moderada para H0	1/10-1/3	Evidência moderada para H1		

*H0: Hipótese nula*

*H1: Hipótese alternativa*

7

8

- **Usar ambos os métodos:** quando selecionado, ambos os métodos de inferência, **Caracterizar distribuição posterior** e **Estimar fator Bayes**, são usados.
4. Selecione e/ou insira as configurações **Variação de dados e Valores de hipótese** apropriadas. A tabela reflete os pares de variáveis que estão atualmente na lista **Variáveis com pares**. À medida que os pares de variáveis são incluídos ou removidos na lista de **Variáveis com pares**, a tabela automaticamente inclui ou remove os mesmos pares de variáveis de suas colunas de pares de variáveis.
- Quando um ou mais pares de variáveis estão na lista **Variáveis pareadas**, as colunas **Variância conhecida** e **Valor de variância** são ativadas.
- Variância conhecida**  
Selecione esta opção para cada variável quando a variação for conhecida.
- Valor da variância**  
Um parâmetro opcional que especifica o valor de variação, se conhecido, para os dados observados.
- Quando um ou mais pares de variáveis estão na lista de **Variáveis com pares** e **Caracterizar distribuição posterior** não é selecionado, as colunas **Valor de teste nulo** e **Valor g** são ativadas.
- Valor de teste nulo**  
Um parâmetro necessário que especifica o valor nulo na estimação do fator de Bayes. Somente um valor é permitido e 0 é o valor padrão.
- Valor de g**  
Especifica o valor a ser definido  $\psi^2 = g\sigma^2_x$  na estimativa do fator bayesiano. Quando o **Valor de variância** é especificado, o **Valor g** é padronizado para 1. Quando o **Valor de variância** não for especificado, é possível especificar um g fixo ou omitir o valor para integrá-lo.
5. Opcionalmente, é possível clicar em **Critérios** para especificar configurações de “Inferência de uma amostra bayesiana: critérios” na página 109 (porcentagem de intervalo crível, opções de valores ausentes e configurações de métodos numéricos) ou clicar em **Anteriores** para especificar configurações de “Inferência de uma amostra bayesiana: anteriores binomiais/de Poisson” na página 113 (distribuições anteriores conjugadas ou customizadas).

## Inferência de amostra independente bayesiana

Este recurso requer a opção SPSS Statistics Standard Edition ou Advanced Statistics.

<sup>7</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>8</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

O procedimento Independente bayesiana - Inferência de amostra fornece opções para utilizar uma variável de grupo para definir dois grupos independentes e executar a inferência bayesiana na diferença das médias dos dois grupos. É possível estimar os fatores bayesianos utilizando abordagens diferentes e também caracterizar a distribuição posterior desejada assumindo que as variações são conhecidas ou desconhecidas.

1. Nos menus, escolha:

**Analisar > Estatísticas bayesianas > Normal de amostras independentes**

2. Selecione as **Variáveis de teste** apropriadas na lista de variáveis de origem. Pelo menos uma variável de origem deve ser selecionada.

3. Selecione a **Variável de agrupamento** apropriada da lista **Variáveis disponíveis**. Uma variável de agrupamento define dois grupos para o teste *t* sem par. A variável de agrupamento selecionada pode ser uma variável numérica ou de sequência.

4. Selecione a **Análise bayesiana**:

- **Caracterizar distribuição posterior**: quando selecionado, a inferência bayesiana é feita de uma perspectiva que é abordada caracterizando as distribuições posteriores. É possível investigar a distribuição marginal posterior dos parâmetros de interesse, integrando os outros parâmetros inconvenientes e construir melhor os intervalos de credibilidade para fazer uma inferência direta. Esta é a definição padrão.
- **Estimar fator Bayes**: quando selecionado, a estimativa de fatores Bayes (uma das metodologias notáveis na inferência bayesiana) constitui uma razão natural para comparar as probabilidades marginais entre um nulo e uma hipótese alternativa.

*Tabela 6. Limites comumente usados para definir a significância de evidência*

Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
>100	Evidência extrema para H0	1-3	Evidência anedótica para H0	1/30-1/10	Evidência forte para H1
30-100	Evidência muito forte para H0	1	Sem evidência	1/100-1/30	Evidência muito forte para H1
10-30	Evidência forte para H0	1/3-1	Evidência anedótica para H1	1/100	Evidência extrema para H1
3-10	Evidência moderada para H0	1/10-1/3	Evidência moderada para H1		

*H0: Hipótese nula*

*H1: Hipótese alternativa*

<sup>9</sup>

<sup>10</sup>

- **Usar ambos os métodos**: quando selecionado, ambos os métodos de inferência, **Caracterizar distribuição posterior** e **Estimar fator Bayes**, são usados.

<sup>9</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>10</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

5. Use as opções **Definir grupos** para definir dois grupos para o teste  $t$ , especificando dois valores (para variáveis de sequência), ou dois valores, um ponto médio ou um ponto de corte (para variáveis numéricas).

**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.

Para variáveis numéricas:

- **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 do ponto médio.** Quando selecionado, os grupos são separados em valores de ponto médio  $<$  e  $\geq$ .
- **Usar 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 que forem menores que o ponto de corte formam um grupo, e os casos com valores que forem 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.

6. Opcionalmente, é possível clicar em **Crêterios** para especificar configurações de “Inferência de amostra independente bayesiana: crêterios” na página 119 (configurações de porcentagem de intervalo de credibilidade, de opções de valores omissos e de método de quadratura adaptativa), clicar em **Parâmetros a priori** para especificar configurações de “Inferência de amostra independente bayesiana: distribuição anterior” na página 119 (variância de dados, a priori em variância e a priori em média condicional em variância) ou clicar em **Estimar fator de Bayes** para especificar configurações de “Bayesiana independente - Inferência de amostra: Estimar fator de Bayes” na página 120.

## Independente bayesiano - Grupos de definição de inferência de amostra (numérico)

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 do ponto médio.** Quando selecionado, os grupos são separados em valores de ponto médio  $<$  e  $\geq$ .
- **Usar 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 que forem menores que o ponto de corte formam um grupo, e os casos com valores que forem maiores ou iguais ao ponto de corte formam o outro grupo.

## Independente bayesiana - Grupos de definição de inferência de amostra (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.

## Inferência de amostra independente bayesiana: critérios

É possível especificar os seguintes critérios de análise para a inferência bayesiana de amostra independente:

### Porcentagem % de intervalo crível

Especifique o nível de significância para os intervalos críveis de cálculo. O nível padrão é 95%.

### Valores omissos

Especifique o método no qual controlar valores omissos.

#### Excluir casos por método pairwise

Esta é a configuração padrão e exclui registros com valores omissos em uma base análise por análise. Os registros que incluem valores omissos, para um campo que é usado por um teste específico, são omitidos do teste.

#### Excluir casos pelo método listwise

Essa configuração exclui os registros que incluem listwise de valores omissos. Os registros que incluem valores omissos para qualquer campo que é nomeado em qualquer subcomando são excluídos de todas as análises.

**Nota:** As opções a seguir estão disponíveis somente quando a opção **Estimar fator bayesiano** ou **Usar ambos os métodos** está selecionada para **Análise bayesiana**.

### Método de quadratura adaptável

Especifique os valores de tolerância e de iteração máxima para o método Quadratura de quadratura adaptável.

#### Tolerância

Especifique o valor de tolerância para os métodos numéricos. A configuração padrão é 0.000001.

#### Máximo de iterações

Especifique o número máximo de iterações do método de Quadratura adaptável. O valor deve ser um número inteiro positivo. A configuração padrão é 500.

## Inferência de amostra independente bayesiana: distribuição anterior

É possível especificar os seguintes critérios de distribuição anteriores para sua Inferência de amostra independente bayesiana:

**Nota:** Muitos pesquisadores aplicados podem questionar a necessidade de especificar a priori. Os a priori de referência minimizam a preocupação de onde o a priori geralmente fica sobrecarregado conforme os dados aumentam. Quando informações de a priori informativo são especificadas, os métodos bayesianos podem efetivamente usar as informações. A necessidade de especificar um a priori não deve ser vista como um impedimento para usar a análise bayesiana.

### Variância de dados

Fornece opções para a definição de configurações de variância de dados.

#### Variância conhecida

Quando selecionado, permite inserir duas variâncias de grupo conhecidas. Ambos os valores devem ser  $> 0$ .

##### Grupo 1 de variância

Insira o valor da primeira variância de grupo conhecida.

##### Variância de grupo 2

Insira o valor da segunda variância de grupo conhecida.

**Assumir variância igual**

Controla se as duas variâncias de grupo devem ser consideradas iguais ou não. Por padrão, supõe-se que as variâncias de grupo sejam desiguais. Esta variável é ignorada quando são inseridos valores para as duas variâncias de grupos.

**Assumir variância desigual**

Controla se as duas variâncias de grupo são consideradas como desiguais ou não. Por padrão, supõe-se que as variâncias de grupo sejam desiguais. Esta variável é ignorada quando são inseridos valores para as duas variâncias de grupos.

**Anterior na variância**

Especifique a distribuição anterior para as duas variâncias iguais.

**Jeffreys**

Quando selecionado, é utilizada uma distribuição anterior não informativa (objetiva) para um espaço de parâmetro.

**Inverse-ChiSquare**

Especifica a distribuição de probabilidade contínua de uma variável aleatória de valor positivo e os parâmetros para  $\chi^2(v_0, \sigma^2_0)$  inverso, em que  $v_0 > 0$  é o grau de liberdade e  $\sigma^2_0 > 0$  é o parâmetro de escala.

**Graus de liberdade**

Especifique um valor para o número de valores no cálculo final que têm liberdade de variação.

**Parâmetro de escala**

Especifique o parâmetro de escala  $\sigma^2_0 > 0$  para  $\chi^2(v_0, \sigma^2_0)$  inverso. Deve-se inserir um valor único maior que 0. Quanto maior o parâmetro de escala, mais difusa a distribuição.

**Anterior na condicional média na variância**

Fornece opções para a especificação da distribuição anterior para as duas médias de grupo.

**Nota:** As opções **Difuso** e **Normal** ficam disponíveis somente quando a opção **Variância conhecida** é selecionada.

**Difuso**

A configuração padrão. Especifica os parâmetros a priori difusos.

**Normal**

Quando selecionado, devem ser especificados parâmetros de local e de escala para as médias de grupo definidas.

**Parâmetro de Localização**

Insira um valor numérico que especifica os parâmetros de localização para as distribuições de grupos.

**Parâmetro de escala**

Especifique o parâmetro de escala  $\sigma^2_0 > 0$  para  $\chi^2(v_0, \sigma^2_0)$  inverso. Para cada grupo, deve-se inserir um valor único que seja maior que 0. Quanto maior o parâmetro de escala, mais difusa a distribuição.

**Bayesiana independente - Inferência de amostra: Estimar fator de Bayes**

É possível especificar o método utilizado para estimar o fator bayesiano.

**Método de Rouder**

Quando selecionado, chama a abordagem Rouder. Esta é a configuração padrão

**Método de Gonen**

Quando selecionado, chama a abordagem de Gonen e as seguintes configurações de tamanho de efeito devem ser especificadas:

**Média para o tamanho do efeito**

Insira um valor que especifica a diferença média entre os dois grupos.

**Variância para o tamanho do efeito**

Insira um valor que especifica a variância para os dois grupos. O valor deve ser  $> 0$ .

### Método Hyperprior

Quando selecionado, chama a abordagem hiper  $g$ , na qual é necessário especificar um valor único. Insira um valor entre -1 e -0,5 no campo **Parâmetro de forma**. O valor padrão é -0,75.

## Inferência bayesiana sobre correlação de Pearson

Este recurso requer a opção SPSS Statistics Standard Edition ou Advanced Statistics.

O coeficiente de correlação de Pearson mede a relação linear entre duas variáveis de escala que seguem juntas uma distribuição normal bivariada. A inferência estatística convencional sobre o coeficiente de correlação foi amplamente discutida e sua prática é oferecida há muito tempo no IBM SPSS Estatísticas. O design da inferência bayesiana sobre o coeficiente de correlação de Pearson permite que os usuários desenhem a inferência bayesiana estimando o fator de Bayes e caracterizando as distribuições posteriores.

1. Nos menus, escolha:

**Analisar > Estatísticas bayesianas > Correlação de Pearson**

2. Selecione as **Variáveis de teste** apropriadas a serem usadas para inferência de correlação pairwise da lista **Variáveis disponíveis**. Pelo menos duas variáveis de origem devem ser selecionadas. Quando mais de duas variáveis são selecionadas, a análise é executada em todas as combinações entre pares das variáveis selecionadas.

3. Selecione a **Análise bayesiana**:

- **Caracterizar distribuição posterior:** quando selecionado, a inferência bayesiana é feita de uma perspectiva que é abordada caracterizando as distribuições posteriores. É possível investigar a distribuição marginal posterior dos parâmetros de interesse, integrando os outros parâmetros inconvenientes e construir melhor os intervalos de credibilidade para fazer uma inferência direta. Esta é a definição padrão.
- **Estimar fator Bayes:** quando selecionado, a estimativa de fatores Bayes (uma das metodologias notáveis na inferência bayesiana) constitui uma razão natural para comparar as probabilidades marginais entre um nulo e uma hipótese alternativa.

Tabela 7. Limites comumente usados para definir a significância de evidência					
Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
>100	Evidência extrema para $H_0$	1-3	Evidência anedótica para $H_0$	1/30-1/10	Evidência forte para $H_1$
30-100	Evidência muito forte para $H_0$	1	Sem evidência	1/100-1/30	Evidência muito forte para $H_1$
10-30	Evidência forte para $H_0$	1/3-1	Evidência anedótica para $H_1$	1/100	Evidência extrema para $H_1$
3-10	Evidência moderada para $H_0$	1/10-1/3	Evidência moderada para $H_1$		

$H_0$ : Hipótese nula

$H_1$ : Hipótese alternativa

11

<sup>11</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

- **Usar ambos os métodos:** quando selecionado, ambos os métodos de inferência, **Caracterizar distribuição posterior** e **Estimar fator Bayes**, são usados.
4. Especifique o **Número máximo de gráficos** a serem exibidos na saída. Um conjunto de gráficos pode conter 3 gráficos na mesma área de janela. Os gráficos são gerados em ordem, começando com a primeira variável em relação às variáveis restantes, depois a segunda variável em relação às variáveis restantes e assim por diante. O valor de número inteiro definido deve estar entre 0 e 50. Por padrão, são gerados 10 conjuntos de gráficos para acomodar cinco variáveis. Esta opção não está disponível quando **Estimar fator de Bayes** é selecionado.
  5. É possível opcionalmente clicar em **Crítérios** para especificar as configurações do “Correlação de Pearson bayesiana: Critérios” na página 122 (porcentagem de intervalo crível, opções de valores omissos e configurações de método numérico), clicar em **Anteriores** para especificar as configurações do “Correlação de Pearson bayesiana: distribuição anterior” na página 123 (valor  $c$  para a anterior  $p(\rho) \propto (1 - \rho^2)^c$  ou clicar em **Fator de Bayes** para especificar as configurações do “Bayesiana independente - Inferência de amostra: Estimar fator de Bayes” na página 120.

## Correlação de Pearson bayesiana: Critérios

É possível especificar os seguintes critérios de análise para a inferência bayesiana de correlação de Pearson (entre pares).

### Porcentagem % de intervalo crível

Especifique o nível de significância para os intervalos críveis de cálculo. O nível padrão é 95%.

### Valores omissos

Especifique o método no qual controlar valores omissos.

#### Excluir casos por método pairwise

Esta configuração exclui os registros que incluem valores omissos entre pares.

#### Excluir casos pelo método listwise

Essa configuração exclui os registros que incluem listwise de valores omissos. Os registros que incluem valores omissos para qualquer campo que é nomeado em qualquer subcomando são excluídos de todas as análises.

**Nota:** As opções a seguir estão disponíveis somente quando a opção **Estimar fator bayesiano** ou **Usar ambos os métodos** está selecionada para **Análise bayesiana**.

### Método Numérico

Especifique o método numérico que é usado para estimar o integral.

#### Configurar valor inicial customizado

Quando selecionada, é possível especificar um valor inicial customizado no campo **Valor inicial**.

#### Tolerância

Especifique o valor de tolerância para os métodos numéricos. A configuração padrão é 0.000001.

#### Máximo de iterações

Especifique o número máximo de iterações de método. O valor deve ser um número inteiro positivo. A configuração padrão é 2000.

#### Número de amostras Monte Carlo

Especifique o número de pontos que servem de amostra para a aproximação de Monte Carlo. O valor deve ser um número inteiro positivo. O valor padrão é 10000.

#### Amostras simuladas para distribuição posterior

Especifique o número de amostras utilizadas para desenhar a distribuição posterior desejada. O valor padrão é 10000.

<sup>12</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.



## Correlação de Pearson bayesiana: distribuição anterior

É possível especificar o valor  $c$  para o anterior  $p(\rho) \propto (1-\rho^2)^c$ .

**Nota:** Muitos pesquisadores aplicados podem questionar a necessidade de especificar a priori. Os a priori de referência minimizam a preocupação de onde o a priori geralmente fica sobrecarregado conforme os dados aumentam. Quando informações de a priori informativo são especificadas, os métodos bayesianos podem efetivamente usar as informações. A necessidade de especificar um a priori não deve ser vista como um impedimento para usar a análise bayesiana.

### Uniforme ( $c = 0$ )

Quando selecionado, o anterior uniforme é utilizado.

### Jeffreys ( $c = -1,5$ )

Quando selecionado, uma distribuição anterior não informativa é utilizada.

### Configurar o valor $c$ customizado

Quando selecionado, é possível especificar um **valor  $c$**  customizado. Qualquer número real único é permitido.

## Correlação de Pearson bayesiana: fator de Bayes

É possível especificar o método utilizado para estimar o fator bayesiano. As opções a seguir estão disponíveis somente quando a opção **Estimar fator bayesiano** ou **Usar ambos os métodos** está selecionada para Análise bayesiana.

### Fator de Bayes JZS

Quando selecionado, chama a abordagem de Zellner-Siow. Esta é a definição padrão.

### Fator de Bayes Fracionária

Quando selecionada, é possível especificar o fator fracionário de Bayes e o valor de hipótese nula. Para o fator fracionário de Bayes, deve-se especificar um valor  $\in (0,1)$ . O valor padrão é 0,5.

## Inferência bayesiana sobre Modelos de regressão linear

Este recurso requer a opção SPSS Statistics Standard Edition ou Advanced Statistics.

A regressão é um método estatístico amplamente utilizado na modelagem quantitativa. A regressão linear é uma abordagem básica e padrão, na qual os pesquisadores usam os valores de diversas variáveis para explicar ou prever valores de um resultado de escala. A regressão linear univariada bayesiana é uma abordagem para a regressão linear na qual a análise estatística é feita no contexto de inferência bayesiana.

É possível chamar o procedimento de regressão e definir um modelo completo.

1. Nos menus, escolha:

**Analisar > Estatísticas bayesianas > Regressão linear**

2. Selecione uma variável dependente única, sem sequência da lista **Variáveis disponíveis**. Deve-se selecionar uma variável sem sequência.
3. Selecione uma ou mais variáveis fator categóricas para o modelo da lista **Variáveis disponíveis**.
4. Selecione uma ou mais variáveis de escala covariável, sem sequência da lista **Variáveis disponíveis**.

**Nota:** As listas **Fator(es)** e **Covariável(is)** não podem ficar vazias. Deve-se selecionar pelo menos uma variável **Fator(es)** ou **Covariável(is)**.

5. Opcionalmente, selecione uma variável única, sem sequência, para servir como a ponderação de regressão da lista **Variáveis disponíveis**.
6. Selecione a **Análise bayesiana**:

- **Caracterizar distribuição posterior:** quando selecionado, a inferência bayesiana é feita de uma perspectiva que é abordada caracterizando as distribuições posteriores. É possível investigar a distribuição marginal posterior dos parâmetros de interesse, integrando os outros parâmetros

inconvenientes e construir melhor os intervalos de credibilidade para fazer uma inferência direta. Esta é a definição padrão.

- **Estimar fator Bayes:** quando selecionado, a estimativa de fatores Bayes (uma das metodologias notáveis na inferência bayesiana) constitui uma razão natural para comparar as probabilidades marginais entre um nulo e uma hipótese alternativa.

*Tabela 8. Limites comumente usados para definir a significância de evidência*

Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
>100	Evidência extrema para H0	1-3	Evidência anedótica para H0	1/30-1/10	Evidência forte para H1
30-100	Evidência muito forte para H0	1	Sem evidência	1/100-1/30	Evidência muito forte para H1
10-30	Evidência forte para H0	1/3-1	Evidência anedótica para H1	1/100	Evidência extrema para H1
3-10	Evidência moderada para H0	1/10-1/3	Evidência moderada para H1		

*H0: Hipótese nula*

*H1: Hipótese alternativa*

13

14

- **Usar ambos os métodos:** quando selecionado, ambos os métodos de inferência, **Caracterizar distribuição posterior** e **Estimar fator Bayes**, são usados.

Como opção, você pode:

- Clique em **Crítérios** para especificar as configurações de porcentagem de intervalo crível e de método numérico.
- Clique em **Anteriores** para definir configurações de distribuição anterior de referência e conjugadas.
- Clique em **Fator bayesiano** para especificar configurações de fator bayesiano.
- Clique em **Salvar** para identificar quais itens devem ser salvos e salvar as informações do modelo em um arquivo XML.
- Clique em **Prever** para especificar regressores para a predição bayesiana.
- Clique em **Gráficos** para plotar as distribuições posteriores dos parâmetros de regressão, a variância de termos de erro e os valores preditos.
- Clique em **Testes F** para comparar modelos estatísticos para identificar o modelo que melhor se ajusta à população da qual provêm os dados de amostra.

## Modelos bayesianos de regressão linear: critérios

É possível especificar os seguintes critérios de análise para os modelos de Regressão linear bayesiana.

<sup>13</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>14</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### Porcentagem % de intervalo crível

Especifique o nível de significância para os intervalos críveis de cálculo. O nível padrão é 95%.

**Nota:** As opções a seguir estão disponíveis somente quando a opção **Estimar fator bayesiano** ou **Usar ambos os métodos** está selecionada para **Análise bayesiana**.

### Método Numérico

Especifique o método numérico que é usado para estimar o integral.

### Tolerância

Especifique o valor de tolerância para os métodos numéricos. A configuração padrão é 0.000001.

### Máximo de iterações

Especifique o número máximo de iterações de método. O valor deve ser um número inteiro positivo. A configuração padrão é 2000.

## Modelos bayesianos de regressão linear: distribuições anteriores

É possível especificar as seguintes configurações de distribuições anteriores para os parâmetros de regressão e a variância de erros. As opções a seguir ficam disponíveis somente quando a opção

**Caracterizar distribuição posterior** é selecionada para **Análise bayesiana**.

**Nota:** Muitos pesquisadores aplicados podem questionar a necessidade de especificar a priori. Os a priori de referência minimizam a preocupação de onde o a priori geralmente fica sobrecarregado conforme os dados aumentam. Quando informações de a priori informativo são especificadas, os métodos bayesianos podem efetivamente usar as informações. A necessidade de especificar um a priori não deve ser vista como um impedimento para usar a análise bayesiana.

### Referência

Quando selecionado, a análise de referência produz a inferência bayesiana objetiva. As instruções de inferência dependem somente do modelo assumido e dos dados disponíveis e a distribuição anterior utilizada para fazer uma inferência é a menos informativa. Esta é a definição padrão.

### Conjugado

Fornece opções para a definição de distribuições anteriores conjugadas. Os anteriores conjugados assumem a distribuição Normal-Inverse-Gamma. Embora conjugados anteriores não sejam necessários ao executar atualizações bayesianas, eles auxiliam no processo de cálculo.

**Nota:** Para especificar parâmetros a priori conjugados para um modelo de regressão linear, configure sua média esperada de parâmetros de regressão na tabela **Parâmetros a priori na variância de erros**. Também é possível optar por usar as configurações de **Variância de matriz de covariâncias** para especificar a variância-covariância a priori.

### Parâmetros a priori na variância de erros

#### Parâmetro de Forma

Especifique o parâmetro de forma  $a_0$  para a distribuição Inverse-Gamma. Deve-se inserir um valor único maior que 0.

#### Parâmetro de escala

Especifique o parâmetro de escala  $b_0$  para a distribuição Inverse-Gamma. Deve-se inserir um valor único maior que 0. Quanto maior o parâmetro de escala, mais difusa a distribuição.

A tabela lista a média de parâmetros de regressão (incluindo o intercepto), que especifica o vetor médio  $\theta_0$  para os parâmetros de regressão definidos. O número de valores deve atender aos parâmetros de regressão, incluindo o termo de intercepto.

O nome da primeira variável é sempre INTERCEPT. Na segunda linha, a coluna **Variáveis** é automaticamente preenchida com as variáveis especificadas por Fator(es) e Covariável(eis). A coluna **Média** não inclui valores padrão.

Clique em **Reconfigurar** para limpar os valores.

### Variância de matriz de covariâncias: $\sigma^2x$

Especifique os valores  $V_0$  no triângulo inferior na matriz de variância-covariância para o anterior normal multivariado. Observe que  $V_0$  deve ser um valor definitivo semi positivo. O último valor de

cada linha deve ser positivo. A linha seguinte deve ter um valor a mais do que a linha anterior. Nenhum valor é especificado para as categorias de referência (se houver).

Clique em **Reconfigurar** para limpar os valores.

#### **Usar matriz de identidade**

Quando selecionado, a matriz de identidade escalada é usada. Não é possível especificar valores  $V_0$  no triângulo inferior na matriz de variância-covariância para o anterior normal multivariado.

## **Modelos bayesianos de regressão linear: fator de Bayes**

É possível especificar o design de modelos para a análise, incluindo a abordagem utilizada para estimar o fator de Bayes para os modelos bayesianos de regressão linear. As seguintes opções estão disponíveis somente quando a opção **Estimar fator de Bayes** ou **Usar ambos os métodos** Análise bayesiana é selecionada.

### **Modelo Nulo**

Quando selecionado, os fatores bayesianos estimados são baseados no modelo nulo. Esta é a definição padrão.

### **Modelo Completo**

Quando selecionado, os fatores bayesianos estimados são baseados no modelo completo e é possível selecionar variáveis a serem usadas e fatores e covariáveis adicionais.

#### **Variáveis**

Lista todas as variáveis disponíveis para o modelo completo.

#### **Fatores adicionais**

Selecione na lista **Variáveis** as variáveis a serem usadas como fatores adicionais.

#### **Covariáveis adicionais**

Selecione na lista **Variáveis** as variáveis a serem usadas como covariáveis adicionais.

### **Cálculo**

Especifique a abordagem para estimar os fatores de Bayes. Método JZS é a configuração padrão.

#### **Método JZS**

Quando selecionado, chama a abordagem de Zellner-Siow. Esta é a definição padrão.

#### **Método de Zellner**

Quando selecionada, chama a abordagem de Zellner e é necessário especificar um valor único  $g$  de a priori  $> 0$  (não há nenhum valor padrão).

#### **Método Hyperprior**

Quando selecionado, chama a abordagem hiper  $g$  e é necessário especificar um parâmetro de forma  $a_0$  para a distribuição gama inversa. Deve-se especificar um valor único  $> 0$  (o valor padrão é 3).

#### **Método de Rouder**

Quando selecionado, chama a abordagem Rouder e é necessário especificar um parâmetro de escala  $b_0$  para a distribuição gama inversa. Deve-se especificar um valor único  $> 0$  (o valor padrão é 1).

## **Modelos bayesianos de regressão linear: Salvar**

Esta caixa de diálogo permite especificar quais estatísticas são pontuadas para a distribuição de predição bayesiana e exportar os resultados do modelo para um arquivo XML.

### **Estatísticas preditivas posteriores**

É possível pontuar as seguintes estatísticas derivadas de predições bayesianas.

#### **Médias**

Média da distribuição preditiva posterior.

#### **Variâncias**

Variância da distribuição preditiva posterior.

### **Modos**

Modo da distribuição preditiva posterior.

### **Limite inferior de intervalo de credibilidade**

Limite inferior do intervalo crível da distribuição preditiva posterior.

### **Limite superior de intervalo de credibilidade**

Limite superior do intervalo crível da distribuição preditiva posterior.

**Nota:** É possível designar nomes de variáveis correspondentes para cada estatística.

### **Exportar informações do modelo para o arquivo XML**

Insira o nome e o local de um arquivo XML para exportar a matriz de variância-covariância do parâmetro pontuado.

## **Modelos bayesianos de regressão linear: prever**

É possível especificar os regressores para gerar distribuições de previsão.

### **Regressores para Predição Bayesiano**

A tabela lista todos os regressores disponíveis. A coluna **Regressores** é automaticamente preenchida com algumas variáveis de Fator e Covariáveis. Especifique os vetores observados com os valores para os regressores. Cada regressor pode receber um valor ou sequência e pode prever somente um caso. Para fatores, são permitidos valores e sequências.

Todos ou nenhum dos valores de regressor devem ser especificados para executar a predição (clicando em **Executar análise**).

Quando uma variável de Fator ou Covariável é removida, a linha do regressor correspondente é removida da tabela.

Para covariáveis, somente valores numéricos podem ser especificados. Para fatores, são permitidos valores numéricos e sequências.

**Nota:** Clique em **Reconfigurar** para limpar os valores definidos.

## **Modelos bayesianos de regressão linear: Gráficos**

É possível controlar os gráficos gerados.

### **Covariáveis**

Lista as covariáveis definidas atualmente.

### **Covariáveis de gráfico**

Na lista **Covariáveis**, selecione as covariáveis a serem incluídas no gráfico e inclua-as na lista **Criar gráfico de covariáveis**.

### **Fatores**

Lista os fatores definidos atualmente.

### **Criar gráfico de fatores**

Na lista **Fatores**, selecione os fatores a serem incluídos no gráfico e inclua-os na lista **Criar gráfico de fatores**.

### **Número máximo de categorias a serem incluídas no gráfico**

Selecione o número máximo de categorias a serem incluídas no gráfico (número inteiro positivo e único). A configuração se aplica a todos os fatores. Por padrão, são incluídos no gráfico os dois primeiros níveis de cada fator.

### **Incluir gráficos de**

#### **Termo Intercepto**

Quando selecionado, o termo de intercepto é incluído no gráfico. A configuração é desmarcada por padrão.

#### **Variância de termos de erro**

Quando selecionado, a variância de erros é expressa no gráfico. A configuração é desmarcada por padrão.

### Distribuição bayesiana predita

Quando selecionado, a distribuição preditiva é incluída no gráfico. A configuração é desmarcada por padrão. A configuração pode ser selecionada apenas quando valores regressores válidos são selecionados.

## Modelos bayesianos de regressão linear: testes F

É possível criar um ou mais testes F parciais. Um teste F é qualquer teste estatístico em que a estatística do teste tem uma distribuição F sob a hipótese nula. Os testes F são comumente utilizados na comparação de modelos estatísticos que foram ajustados a um conjunto de dados, a fim de identificar o modelo que melhor se ajusta à população da qual provêm os dados de amostra.

### Variáveis disponíveis

Lista as variáveis de fator e covariáveis selecionadas na caixa de diálogo Regressão linear bayesiana principal. Quando as variáveis fator e covariáveis são incluídas ou removidas do diálogo principal, a lista é atualizada de forma apropriada.

### Variáveis de teste

Selecione as variáveis fator/covariável para testar da lista **Variáveis disponíveis** e inclua-as na lista **Variáveis de teste**.

**Nota:** A opção **Incluir termo de intercepto** deve ser selecionada quando nenhum fator ou covariável de teste estiver selecionada.

### Variáveis e valores de teste

Especifique os valores a serem testados. O número de valores deve corresponder ao número de parâmetros no modelo original. Quando forem especificados valores, o primeiro valor deve ser especificado para o termo de intercepto (suponha que todos os valores são 0 quando não forem definidos explicitamente).

### Incluir termo de intercepto

Quando selecionado, os termos de intercepto são incluídos no teste. Por padrão, a configuração não é selecionada.

Quando ativado, use o campo **Valor de teste** para especificar um valor.

### Rótulo de Teste (opcional)

Opcionalmente, é possível especificar um rótulo para cada teste. É possível especificar um valor de sequência com comprimento máximo de 255 bytes. É permitido somente um rótulo para cada teste F.

## ANOVA Bayesiano de um fator

Este recurso requer a opção SPSS Statistics Standard Edition ou Advanced Statistics.

O procedimento ANOVA de um fator produz uma análise de variância por um fator para uma variável dependente quantitativa por uma única variável (independente) de fator. A análise de variância é utilizada para testar a hipótese de que várias médias são iguais. O SPSS Statistics suporta fatores bayesianos, anteriores conjugados e anteriores não informativos.

1. Nos menus, escolha:

**Analisar > Estatísticas bayesianas > ANOVA de um fator**

2. Selecione uma variável **Dependente** única, numérica da lista **Variáveis disponíveis**. Deve-se selecionar pelo menos uma variável.
3. Selecione uma variável **Fator** única para o modelo da lista **Variáveis disponíveis**. Deve-se selecionar pelo menos uma variável **Fator**.
4. Selecione uma variável única, sem sequência para servir como a **Ponderação** de regressão da lista **Variáveis disponíveis**. O campo de variável **Ponderação** pode ficar vazio.
5. Selecione a **Análise bayesiana**:
  - **Caracterizar distribuição posterior:** quando selecionado, a inferência bayesiana é feita de uma perspectiva que é abordada caracterizando as distribuições posteriores. É possível investigar a

distribuição marginal posterior dos parâmetros de interesse, integrando os outros parâmetros inconvenientes e construir melhor os intervalos de credibilidade para fazer uma inferência direta. Esta é a definição padrão.

- **Estimar fator Bayes:** quando selecionado, a estimativa de fatores Bayes (uma das metodologias notáveis na inferência bayesiana) constitui uma razão natural para comparar as probabilidades marginais entre um nulo e uma hipótese alternativa.

*Tabela 9. Limites comumente usados para definir a significância de evidência*

Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
>100	Evidência extrema para H0	1-3	Evidência anedótica para H0	1/30-1/10	Evidência forte para H1
30-100	Evidência muito forte para H0	1	Sem evidência	1/100-1/30	Evidência muito forte para H1
10-30	Evidência forte para H0	1/3-1	Evidência anedótica para H1	1/100	Evidência extrema para H1
3-10	Evidência moderada para H0	1/10-1/3	Evidência moderada para H1		

*H0: Hipótese nula*

*H1: Hipótese alternativa*

15

16

- **Usar ambos os métodos:** quando selecionado, ambos os métodos de inferência, **Caracterizar distribuição posterior** e **Estimar fator Bayes**, são usados.

Como opção, você pode:

- Clique em **Crítérios** para especificar as configurações de porcentagem de intervalo crível e de método numérico.
- Clique em **Anteriores** para definir configurações de distribuição anterior de referência e conjugadas.
- Clique em **Fator bayesiano** para especificar configurações de fator bayesiano.
- Clique em **Gráficos** para controlar os gráficos gerados.

## ANOVA bayesiano unidirecional: critérios

É possível especificar os critérios de análise a seguir para modelos ANOVA bayesianos unidirecionais.

### Porcentagem % de intervalo crível

Especifique o nível de significância para os intervalos críveis de cálculo. O nível padrão é 95%.

**Nota:** As opções a seguir estão disponíveis somente quando a opção **Estimar fator bayesiano** ou **Usar ambos os métodos** está selecionada para **Análise bayesiana**.

### Método Numérico

Especifique o método numérico que é usado para estimar o integral.

<sup>15</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>16</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

**Tolerância**

Especifique o valor de tolerância para os métodos numéricos. A configuração padrão é 0.000001.

**Máximo de iterações**

Especifique o número máximo de iterações de método. O valor deve ser um número inteiro positivo. A configuração padrão é 2000.

**ANOVA bayesiano de um fator: anteriores**

É possível especificar as seguintes configurações de distribuições anteriores para os parâmetros de regressão e a variância de erros. As opções a seguir ficam disponíveis somente quando a opção

**Caracterizar distribuição posterior** é selecionada para **Análise bayesiana**.

**Nota:** Muitos pesquisadores aplicados podem questionar a necessidade de especificar a priori. Os a priori de referência minimizam a preocupação de onde o a priori geralmente fica sobrecarregado conforme os dados aumentam. Quando informações de a priori informativo são especificadas, os métodos bayesianos podem efetivamente usar as informações. A necessidade de especificar um a priori não deve ser vista como um impedimento para usar a análise bayesiana.

**Referência**

Quando selecionado, a análise de referência produz a inferência bayesiana objetiva. As instruções de inferência dependem somente do modelo assumido e dos dados disponíveis e a distribuição anterior utilizada para fazer uma inferência é a menos informativa. Esta é a definição padrão.

**Conjugado**

Fornece opções para a definição de distribuições anteriores conjugadas. Os anteriores conjugados assumem a distribuição Normal-Inverse-Gamma. Embora conjugados anteriores não sejam necessários ao executar atualizações bayesianas, eles auxiliam no processo de cálculo.

**Parâmetros a priori na variância de erros****Parâmetro de Forma**

Especifique o parâmetro de forma  $a_0$  para a distribuição Inverse-Gamma. Deve-se inserir um valor único maior que 0.

**Parâmetro de escala**

Especifique o parâmetro de escala  $b_0$  para a distribuição Inverse-Gamma. Deve-se inserir um valor único maior que 0. Quanto maior o parâmetro de escala, mais difusa a distribuição.

**Anteriores em parâmetros de regressão**

Especifique o vetor médio  $\beta_0$  para as médias de grupo. O número de valores deve atender aos parâmetros de regressão, incluindo o termo de intercepto.

A coluna **Variáveis** é preenchida automaticamente com os níveis do fator. A coluna **Média** não inclui valores padrão.

Clique em **Reconfigurar** para limpar os valores.

**Variância de matriz de covariâncias:  $\sigma^2x$** 

Especifique os valores  $V_0$  no triângulo inferior na matriz de variância-covariância para o anterior normal multivariado. Observe que  $V_0$  deve ser um valor definitivo semi positivo. Somente o triângulo inferior da tabela deve ser especificado.

As linhas e colunas são automaticamente preenchidas com os níveis do fator. Todos os valores diagonais são 1; todos os valores fora da diagonal são 0.

Clique em **Reconfigurar** para limpar os valores.

**Usar matriz de identidade**

Quando selecionado, a matriz de identidade é usada. Não é possível especificar valores  $V_0$  no triângulo inferior na matriz de variância-covariância para o anterior normal multivariado.



## ANOVA bayesiano de um fator: fator de Bayes

É possível especificar a abordagem utilizada para estimar o fator de Bayes para os modelos ANOVA bayesianos de um fator. As seguintes opções estão disponíveis somente quando a opção **Estimar fator de Bayes** ou **Usar ambos os métodos** Análise bayesiana é selecionada.

### Cálculo

Especifique a abordagem para estimar os fatores de Bayes. Método JZS é a configuração padrão.

#### Método JZS

Quando selecionado, chama a abordagem de Zellner-Siow. Esta é a definição padrão.

#### Método de Zellner

Quando selecionada, chama a abordagem de Zellner e é necessário especificar um valor único  $g$  de a priori  $> 0$  (não há nenhum valor padrão).

#### Método Hyperprior

Quando selecionado, chama a abordagem hiper  $g$  e é necessário especificar um parâmetro de forma  $\alpha_0$  para a distribuição gama inversa. Deve-se especificar um valor único  $> 0$  (o valor padrão é 3).

#### Método de Rouder

Quando selecionado, chama a abordagem Rouder e é necessário especificar um parâmetro de escala  $b_0$  para a distribuição gama inversa. Deve-se especificar um valor único  $> 0$  (o valor padrão é 1).

## ANOVA bayesiano unidirecional: gráficos

É possível controlar os gráficos gerados.

### Criar gráfico do(s) grupo(s)

Especifique os subgrupos dos quais será criado um gráfico. Crie um gráfico com a probabilidade, o anterior e o posterior para as médias dos grupos especificados. A lista **Grupos** é um subconjunto das categorias da variável fator, portanto, o formato deve ser consistente com o tipo de dados do fator e com os valores reais.

### Variância de termos de erro

Quando selecionado, a variância de erros é expressa no gráfico. A configuração é desmarcada por padrão. Esta opção não fica disponível quando **Estimar fator de Bayes** é selecionado como a Análise bayesiana.

## Modelos Log-linear bayesianos

Este recurso requer a opção SPSS Statistics Standard Edition ou Advanced Statistics.

O projeto para testar a independência de dois fatores requer duas variáveis categóricas para a construção de uma tabela de contingência e faz a inferência bayesiana na associação entre linha e coluna. É possível estimar os fatores bayesianos assumindo diferentes modelos e caracterizar a distribuição posterior desejada simulando o intervalo crível simultâneo para os termos de interação.

1. Nos menus, escolha:

#### **Analisar > Estatísticas bayesianas > Modelos log-lineares**

2. Selecione uma variável de linha única, sem escala da lista **Variáveis disponíveis**. Deve-se selecionar pelo menos uma variável sem escala.
3. Selecione uma variável da coluna única, sem escala da lista **Variáveis disponíveis**. Deve-se selecionar pelo menos uma variável sem escala.
4. Selecione a **Análise bayesiana**:
  - **Caracterizar distribuição posterior**: quando selecionado, a inferência bayesiana é feita de uma perspectiva que é abordada caracterizando as distribuições posteriores. É possível investigar a distribuição marginal posterior dos parâmetros de interesse, integrando os outros parâmetros

inconvenientes e construir melhor os intervalos de credibilidade para fazer uma inferência direta. Esta é a definição padrão.

- **Estimar fator Bayes:** quando selecionado, a estimativa de fatores Bayes (uma das metodologias notáveis na inferência bayesiana) constitui uma razão natural para comparar as probabilidades marginais entre um nulo e uma hipótese alternativa.

Tabela 10. Limites comumente usados para definir a significância de evidência					
Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
>100	Evidência extrema para H0	1-3	Evidência anedótica para H0	1/30-1/10	Evidência forte para H1
30-100	Evidência muito forte para H0	1	Sem evidência	1/100-1/30	Evidência muito forte para H1
10-30	Evidência forte para H0	1/3-1	Evidência anedótica para H1	1/100	Evidência extrema para H1
3-10	Evidência moderada para H0	1/10-1/3	Evidência moderada para H1		

*H0: Hipótese nula*

*H1: Hipótese alternativa*

17

18

- **Usar ambos os métodos:** quando selecionado, ambos os métodos de inferência, **Caracterizar distribuição posterior** e **Estimar fator Bayes**, são usados.

Como opção, você pode:

- Clique em **Crítérios** para especificar as configurações de porcentagem de intervalo crível e de método numérico.
- Clique em **Fator bayesiano** para especificar configurações de fator bayesiano.
- Clique em **Imprimir** e especifique como o conteúdo é exibido nas tabelas de saída.

## Modelos log-linear bayesianos: Critérios

É possível especificar os seguintes critérios de análise para modelos log-linear bayesianos.

### Porcentagem % de intervalo crível

Especifique o nível de significância para os intervalos críveis de cálculo. O nível padrão é 95%.

### Método Numérico

Especifique o método numérico que é usado para estimar o integral.

### Configurar valor inicial customizado

Quando selecionada, é possível especificar um valor inicial customizado no campo **Valor inicial**. Especifique um valor inicial aleatório configurado. O valor deve ser um número inteiro positivo. Por padrão, um valor inicial aleatório é designado.

<sup>17</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>18</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

**Nota:** As opções a seguir estão disponíveis somente quando a opção **Estimar fator bayesiano** ou **Usar ambos os métodos** está selecionada para **Análise bayesiana**.

#### **Tolerância**

Especifique o valor de tolerância para os métodos numéricos. A configuração padrão é 0.000001.

#### **Máximo de iterações**

Especifique o número máximo de iterações de método. O valor deve ser um número inteiro positivo. A configuração padrão é 2000.

#### **Amostras simuladas para a Distribuição posterior**

Especifique o número de amostras utilizadas para desenhar a distribuição posterior desejada. O valor padrão é 10000.

#### **Formato**

Selecione se as categorias devem ser exibidas em ordem **Crescente** ou **Decrescente**. Crescente é a configuração padrão.

## **Modelos log-linear bayesianos: Fator de Bayes**

É possível especificar o modelo assumido para os dados observados (Poisson, multinomial ou não paramétrico). A distribuição multinomial é a configuração padrão. As seguintes opções estão disponíveis somente quando a opção **Estimar fator de Bayes** ou **Usar ambos os métodos** Análise bayesiana é selecionada.

#### **Modelo de Poisson**

Quando selecionado, o modelo de Poisson é assumido para os dados observados.

#### **Modelo multinomial**

Quando selecionado, o modelo multinomial é assumido para os dados observados. Esta é a definição padrão.

#### **Margens Fixas**

Selecione **Total geral**, **Soma da linha** ou **Soma da coluna** para especificar os totais marginais fixos para a tabela de contingência. **Total geral** é a configuração padrão.

#### **Distribuição Anterior**

Especifique o tipo de distribuição anterior ao estimar o fator de Bayes.

##### **Conjugado**

Selecione para especificar uma distribuição anterior conjugada. Use a tabela **Parâmetros de forma** para especificar os parâmetros de forma  $a_{rs}$  para a distribuição Gama. Os parâmetros de forma devem ser especificados quando **Conjugado** for selecionado como o tipo de distribuição anterior.

Quando um único valor é especificado, todos os  $a_{rs}$  são considerados iguais a esse valor.  $a_{rs} = 1$  é a configuração padrão. Se for necessário especificar mais de um valor, é possível separar os valores com espaços em branco.

O número de valores numéricos especificados em cada linha e cada coluna deve corresponder à dimensão da tabela de contingência. Todos os valores especificados devem ser  $> 0$ .

Clique em **Reconfigurar** para limpar os valores.

##### **Parâmetro de escala**

Especifique o parâmetro de escala  $b$  para distribuição Gama. Deve-se especificar um valor único  $> 0$ .

#### **Mistura Dirichlet**

Selecione para especificar uma distribuição anterior Dirichlet mista.

#### **Intrínseco**

Selecione para especificar uma distribuição anterior intrínseca.

#### **Modelo Paramétrico**

Quando selecionado, o modelo não paramétrico é assumido para os dados observados.

### Margens Fixas

Selecione **Soma da linha** ou **Soma da coluna** para especificar os totais marginais fixos para a tabela de contingência. **Soma da linha** é a configuração padrão.

### Distribuição Anterior

Especifique os parâmetros para os anteriores de Dirichlet. Deve-se especificar os parâmetros de **Distribuição anterior** quando o **Modelo não paramétrico** for selecionado. Quando um único valor é especificado, todos os  $\lambda_s$  são considerados iguais a esse valor.  $\lambda_s = 1$  é a configuração padrão. Se for necessário especificar mais de um valor, é possível separar os valores com espaços em branco. Todos os valores especificados devem ser  $> 0$ . O número de valores numéricos especificados deve corresponder à dimensão da linha ou coluna que não é fixa para a tabela de contingência.

Clique em **Reconfigurar** para limpar os valores.

## Modelos log-linear bayesianos: Impressão

É possível especificar como o conteúdo deve ser exibido nas tabelas de saída.

### Design da Tabela

#### Suprimir Tabela

Quando selecionado, a tabela de contingência não é incluída na saída. A configuração não é ativada por padrão.

**Nota:** As configurações a seguir não têm efeito quando a configuração **Suprimir tabela** está ativada.

### Estatísticas

Especifique as estatísticas para testar a independência.

#### Qui-quadrado

Selecione para calcular a estatística qui-quadrado de Pearson, os graus de liberdade e a significância assintótica de dois lados. Para uma tabela de contingência de 2 por 2, essa configuração também calcula as estatísticas corrigidas de continuidade de Yates, os graus de liberdade e a significância assintótica de dois lados associada. Para uma tabela de contingência de 2 por 2, com pelo menos uma contagem de células esperada  $< 5$ , essa configuração também calcula a significância exata de dois lados e de um lado do teste exato de Fisher.

#### Razão de verossimilhança

Selecione para calcular a estatística de teste da razão de verossimilhança, os graus de liberdade e a significância assintótica de dois lados associada.

### Contagens

Especifique quais tipos de contagem são incluídos na tabela de contingência.

#### Observado

Selecione para incluir as contagens de células observadas na tabela de contingência.

#### Esperado

Selecione para incluir as contagens de células esperadas na tabela de contingência.

### Porcentagens

Especifique quais tipos de porcentagem são incluídos na tabela de contingência.

#### Linha

Selecione para incluir porcentagens de linha na tabela de contingência.

#### Coluna

Selecione para incluir porcentagens de coluna na tabela de contingência.

#### Completa

Selecione para incluir porcentagens totais na tabela de contingência.

## Modelos de ANOVA de medidas repetidas de um fator bayesiana

Este recurso requer a opção SPSS Statistics Standard Edition ou Advanced Statistics.

Em modelos de análise de variância para um fator (ANOVA) bayesiana, supõe-se que haja uma única medida por assunto. No entanto, essa suposição nem sempre é verdadeira. Não é incomum que um design de estudo seja destinado a investigar médias de respostas em vários momentos ou condições. O procedimento ANOVA de medidas repetidas de um fator bayesiana mede um fator do mesmo assunto a cada momento ou condição distinta e permite que os assuntos sejam cruzados dentro dos níveis. Supõe-se que cada assunto tenha uma única observação para cada momento ou condição (portanto, a interação de assunto-tratamento não é considerada).

1. Nos menus, escolha:

**Analisar > Estatísticas bayesianas > ANOVA de medidas repetidas de um fator**

2. Selecione pelo menos duas variáveis de **Medidas repetidas** da lista **Variáveis disponíveis**.

3. Opcionalmente, selecione uma única variável para servir como a **Ponderação** de regressão da lista **Variáveis disponíveis**. O campo de variável **Ponderação** pode ficar vazio.

**Nota:** A lista de variáveis disponíveis fornece todas as variáveis, exceto as variáveis de Sequência.

4. Selecione a **Análise bayesiana**:

- **Caracterizar distribuição posterior:** quando selecionado, a inferência bayesiana é feita de uma perspectiva que é abordada caracterizando as distribuições posteriores. É possível investigar a distribuição marginal posterior dos parâmetros de interesse, integrando os outros parâmetros inconvenientes e construir melhor os intervalos de credibilidade para fazer uma inferência direta. Esta é a definição padrão.
- **Estimar fator Bayes:** quando selecionado, a estimativa de fatores Bayes (uma das metodologias notáveis na inferência bayesiana) constitui uma razão natural para comparar as probabilidades marginais entre um nulo e uma hipótese alternativa.

*Tabela 11. Limites comumente usados para definir a significância de evidência*

Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência	Fator de Bayes	Categoria de evidência
>100	Evidência extrema para H0	1-3	Evidência anedótica para H0	1/30-1/10	Evidência forte para H1
30-100	Evidência muito forte para H0	1	Sem evidência	1/100-1/30	Evidência muito forte para H1
10-30	Evidência forte para H0	1/3-1	Evidência anedótica para H1	1/100	Evidência extrema para H1
3-10	Evidência moderada para H0	1/10-1/3	Evidência moderada para H1		

*H0: Hipótese nula*

*H1: Hipótese alternativa*

19

20

- **Usar ambos os métodos:** quando selecionado, ambos os métodos de inferência, **Caracterizar distribuição posterior** e **Estimar fator Bayes**, são usados.

<sup>19</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>20</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Como opção, você pode:

- Clique em **Crítérios** para especificar as configurações de porcentagem de intervalo crível e de método numérico.
- Clique em **Fator bayesiano** para especificar configurações de fator bayesiano.
- Clique em **Gráficos** para plotar as distribuições posteriores das médias do grupo.

## ANOVA de medidas repetidas de um fator bayesiana: Critérios

É possível especificar os seguintes critérios de análise para modelos de ANOVA de medidas repetidas de um fator bayesiana.

### Porcentagem % de intervalo crível

Especifique o nível de significância para os intervalos críveis de cálculo. O nível padrão é 95%.

### Método Numérico

Especifique o método numérico que é usado para estimar o integral.

### Configurar valor inicial customizado

Quando selecionada, é possível especificar um valor inicial customizado no campo **Valor inicial**. O valor padrão é 2.000.000. O valor deve ser um número inteiro positivo entre 1 e 2.147.483.647. Por padrão, um valor inicial aleatório é designado.

### Número de amostras Monte Carlo

Especifique o número de pontos que servem de amostra para a aproximação de Monte Carlo. O valor deve ser um número inteiro positivo entre  $10^3$  e  $10^6$ . O valor padrão é 30.000.

## ANOVA de medidas repetidas de um fator bayesiana: Fator de Bayes

É possível especificar a abordagem que é usada para estimar o fator de Bayes para os modelos de ANOVA de medidas repetidas de um fator bayesiana. As seguintes opções estão disponíveis somente quando a opção **Estimar fator de Bayes** ou **Usar ambos os métodos** Análise bayesiana é selecionada.

### Crítérios de informações bayesianas (BIC)

Usa uma extensão da aproximação BIC para o design de medidas repetidas para estimar fatores de Bayes. A configuração deriva a contabilidade de tamanho da amostra efetivo para correlação de medidas repetidas, e sugere um termo de penalidade melhorado ao estimar o BIC para a seleção entre dois modelos concorrentes. Esta é a definição padrão.

### Design misto de Roudier

Usa as generalizações multivariadas da distribuição Cauchy como a priori para tamanho do efeito padronizado e um a priori não informativo para variância.

**Nota:** A configuração de ponderação de frequência global e a ponderação de regressão são ignoradas quando esta opção é selecionada.

## ANOVA de medidas repetidas de um fator bayesiana: Gráficos

É possível controlar os gráficos que são exibidos na saída para ilustrar as distribuições posteriores das médias do grupo. A tabela lista todas as variáveis que são selecionadas como medidas repetidas no diálogo Variáveis. Selecione as variáveis de medida repetida para o gráfico

## Regressão de Kernel Ridge

A Regressão de Kernel Ridge é um procedimento de extensão que usa a classe Python **sklearn.kernel\_ridge.KernelRidge** para estimar modelos de regressão de kernel ridge estimados. Os modelos de regressão de kernel ridge são modelos de regressão não paramétricos capazes de modelar relacionamentos lineares e não lineares entre variáveis preditoras e resultados. Os resultados podem ser altamente sensíveis às escolhas de hiperparâmetros de modelo. A Regressão de Kernel Ridge facilita a escolha de valores de hiperparâmetro por meio de validação cruzada de dobra k em grades especificadas de valores usando a classe **sklearn.model\_selection.GridSearchCV**.

### Exemplo:

#### Estatísticas

Additive\_CHI2, CHI2, Cosseno, Laplaciano, Linear, Polinomial, RBF, Sigmoide, Alfa, Gama, Coef0, Grau, validação cruzada, observada versus predita, resíduos versus predito, coeficientes de ponderação dupla, coeficientes de ponderação de espaço do kernel.

## Considerações de dados

### Dados

- É possível especificar qualquer ou todas as oito diferentes funções kernel.
- A função kernel selecionada determina quais hiperparâmetros estão ativos.
- Os hiperparâmetros incluem alfa para regularização de ridge comuns a todos os kernels mais tantos quantos três outros hiperparâmetros para cada função kernel específica.
- Quando vários subcomandos de kernel são especificados, ou mais de um valor para qualquer parâmetro é especificado, uma procura de grade com validação cruzada para avaliar modelos é executada, e o modelo com melhor ajuste com base nos dados mantidos é selecionado.
- A extensão aceita variáveis de divisão do procedimento de Arquivo de Divisão e ponderações usando o procedimento de Casos de Ponderação.
- Quando as ponderações são incluídas, elas são usadas para criar valores ajustados em todas as análises. Devido a limitações no método de escoragem na classe **sklearn.model\_selection.GridSearchCV**, as avaliações de validação cruzada usadas para a seleção de modelo não são ponderadas.

### Suposições

## Obtendo uma Regressão de Kernel Ridge

1. Nos menus, escolha:

**Analisar > Regressão > Kernel Ridge...**

2. Selecione uma variável **Dependente**.
3. Selecione uma ou mais variáveis **Independente**.
4. A configuração padrão **Modelo único** é usada quando apenas um valor para cada parâmetro de função kernel é especificado. Quando a configuração **Modelo único** é selecionada, não é possível especificar funções **Kernel** adicionais e ponderações são totalmente aplicadas em toda a análise, avaliação e pontuação de resultados. Também é possível usar os controles de seta para cima e para baixo para reorganizar as funções kernel.

Opcionalmente, selecione **Seleção de modelo** da lista **Modo**.

Quando **Seleção de modelo** é selecionado da lista **Modo**, é possível incluir múltiplas funções kernel à lista de **Kernels**.

- a. Clique no controle de inclusão (+) para incluir funções adicionais do kernel.
- b. Clique na célula vazia na coluna **Kernel** para selecionar uma função kernel.
- c. Clique duas vezes em qualquer célula de linha de função kernel para especificar valores de parâmetro de função kernel para a coluna correspondente (**Alfa**, **Gamma**, **Coef0**, **Grau**). Para obter informações adicionais, consulte “Parâmetros de kernel” na página 138. Os parâmetros de ajuste da função kernel padrão são listados abaixo.

#### **Additive\_CHI2**

ALPHA=1 GAMMA=1

#### **CHI2**

ALPHA=1 GAMMA=1

#### **Cosseno**

ALPHA=1

**Laplaciano**

ALPHA=1 GAMMA=1/p

**Linear**

A função kernel padrão. ALPHA=1

**Polinomial**

ALPHA=1 GAMMA=1/p COEF0=1 DEGREE=3

**RBF**

ALPHA=1 GAMMA=1/p

**Curva sigmoide**

ALPHA=1 GAMMA=1/p COEF0=1

**Nota:** Quando mais de um valor para qualquer parâmetro de função kernel é especificado, uma procura de grade com validação cruzada para avaliar modelos é executada e o modelo com melhor ajuste baseado nos dados mantidos é selecionado.

5. Opcionalmente, clique em **Opções** para especificar o número de dobras de validação cruzada, opções de exibição, configurações de gráfico e itens a serem salvos. Para obter mais informações, consulte [“Regressão de Kernel Ridge: opções”](#) na página 138.
6. Clique em **OK**.

## Parâmetros de kernel

O diálogo **Parâmetros de Kernel** fornece opções para especificar valores de parâmetro de função de kernel único e para especificar que aquela seleção de modelo seja executada usando uma procura de grade sobre as combinações de kernels e valores de parâmetro de grade especificados.

**Especificar parâmetros únicos**

Ative a configuração para especificar valores para o parâmetro de função do kernel selecionado.

- Insira um valor e clique em **Incluir** para incluir o valor no parâmetro de função do kernel.
- Selecione um valor de parâmetro e clique em **Alterar** para atualizar o valor.
- Selecione um valor de parâmetro e clique em **Remover** para excluir o valor.

**Especificar parâmetros de grade**

Ative a configuração para especificar que a seleção do modelo seja executada usando uma procura de grade nas combinações de kernels e valores de parâmetro de grade especificados.

## Regressão de Kernel Ridge: opções

O diálogo **Gráficos** fornece opções para especificar o número de dobras de validação cruzada, opções de exibição, configurações de gráfico e itens a serem salvos.

**Número de pastas de validação cruzada**

O número de divisões ou dobras na validação cruzada com procura de grade para seleção de modelo. Insira um valor de número inteiro maior que 1. O valor padrão é 5 segundos. A configuração está disponível apenas quando **Seleção de modelo** é escolhida como o **Modo** no diálogo primário **Regressão de Kernel Ridge**.

**Exibir**

Fornece opções para especificar qual saída exibir quando a validação cruzada estiver em vigor.

**Melhor**

A configuração padrão exibe apenas resultados básicos para o melhor modelo escolhido.

**Comparar**

Exibe resultados básicos para todos os modelos avaliados.

**Comparar modelos e pastas**

Exibe resultados completos para cada divisão ou dobra para cada modelo avaliado.



**Gráfico**

Fornecer opções para especificar gráficos de valores observados ou residuais em relação a valores preditos.

**Observados vs. preditos**

Exibe um gráfico de dispersão de valores observados em relação aos preditos para o modelo especificado ou melhor.

**Resíduos vs. preditos**

Exibe um gráfico de dispersão de valores residuais em relação aos preditos para o modelo especificado ou melhor.

**Salvar**

A tabela fornece opções para especificar variáveis para salvar no conjunto de dados ativo.

**Valores preditos**

Salva valores preditos do modelo especificado ou melhor no conjunto de dados ativo. Um nome de variável opcional pode ser incluído.

**Resíduos**

Salva resíduos das predições do modelo especificado ou melhor no conjunto de dados ativo. Um nome de variável opcional pode ser incluído.

**Coeficientes dual**

Salva os coeficientes de ponderação de espaço duplo ou de kernel do modelo especificado no conjunto de dados ativo. Um nome de variável opcional pode ser incluído. A configuração não está disponível quando **Seleção de modelo** é escolhida como o **Modo** no diálogo primário **Regressão de Kernel Ridge**.



## Avisos

---

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

É possível que a IBM não ofereça 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 produtos, programas ou serviços IBM não significa que apenas produtos, programas ou serviços IBM possam ser utilizados. Qualquer outro produto, programa ou serviço, funcionalmente equivalente, poderá ser utilizado em substituição daqueles, desde que não infrinja nenhum direito de propriedade intelectual da IBM. Entretanto, a avaliação e verificação da operação de qualquer produto, programa ou serviço não IBM são de responsabilidade do Cliente.

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

*Gerência de Relações Comerciais e Industriais da IBM Brasil*  
*IBM Brasil Ltda*  
*Botafogo*  
*Rio de Janeiro, RJ*  
*CEP 22290-240*

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

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

A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO "NO ESTADO EM QUE SE ENCONTRA", SEM GARANTIA DE NENHUM TIPO, SEJA EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS 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. São feitas alterações periódicas nas informações aqui contidas; tais alterações serão incorporadas em futuras edições desta publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar os produtos e/ou programas descritos nesta publicação, sem aviso prévio.

Todas as referências nestas informações para websites não IBM são fornecidas apenas para conveniência e não de forma a servir como 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:

*Gerência de Relações Comerciais e Industriais da IBM Brasil*  
*IBM Brasil Ltda*

*Botafogo  
Rio de Janeiro, RJ  
CEP 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. Resultados de desempenho reais podem variar dependendo das configurações específicas e das condições operacionais.

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

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

Estas informações contêm exemplos de dados e relatórios utilizados nas operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos 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 amostra sem a necessidade de pagar à IBM, com objetivos de desenvolvimento, utilização, marketing ou distribuição de programas aplicativos em conformidade com a interface de programação de aplicativo para a plataforma operacional para a qual os programas de amostra são criados. Esses exemplos não foram testados completamente em todas as condições. Portanto, a IBM não pode garantir ou implicar a confiabilidade, manutenção ou função destes programas. Os programas de amostra são fornecidos "NO ESTADO EM QUE SE ENCONTRAM", sem garantia de qualquer tipo. A IBM não será responsabilizada por quaisquer danos decorrentes do uso dos programas de amostra.

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

© 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 todo. Outros nomes de produto e de serviço 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 registradas ou marcas comerciais da Adobe Systems Incorporated nos Estados Unidos e/ou em outros países.

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

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

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

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

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



# Índice remissivo

## A

amplitude múltipla de Ryan-Einot-Gabriel-Welsch  
  no GLM Multivariate [5](#)  
  no GLM Repeated Measures [14](#)  
análise de covariância  
  no GLM Multivariate [1](#)  
Análise de log linear de seleção do modelo  
  definindo intervalos de fator [65](#)  
  modelos [66](#)  
  opções [66](#)  
  recursos adicionais do comando [66](#)  
Análise de log linear geral  
  contrastes [67](#)  
  covariáveis de célula [67](#)  
  critérios [68](#)  
  distribuição de contagens de células [67](#)  
  especificação de modelo [68](#)  
  estruturas de célula [67](#)  
  fatores [67](#)  
  gráficos [68](#)  
  intervalos de confiança [68](#)  
  opções de exibição [68](#)  
  recursos adicionais do comando [69](#)  
  resíduos [69](#)  
  salvando valores preditos [69](#)  
  salvando variáveis [69](#)  
análise de probit  
  modelos lineares generalizados mistos [50](#)  
análise de sobrevivência  
  em Kaplan-Meier [92](#)  
  em Regressão de Cox [94](#)  
  em regressão de Kernel Ridge [136](#)  
  em Tabelas de Mortalidade [72](#)  
  Regressão de Cox dependente de tempo [98](#)  
análise de variância  
  em Componentes de Variância [20](#)  
  em modelos lineares generalizados mistos [50](#)  
Análise Log-Linear Logit  
  contrastes [69](#)  
  covariáveis de célula [69](#)  
  critérios [71](#)  
  distribuição de contagens de células [69](#)  
  especificação de modelo [71](#)  
  estruturas de célula [69](#)  
  fatores [69](#)  
  gráficos [71](#)  
  intervalos de confiança [71](#)  
  opções de exibição [71](#)  
  resíduos [72](#)  
  salvando variáveis [72](#)  
  valores preditos [72](#)  
análise loglinear  
  Análise de log linear geral [67](#)  
  Análise Log-Linear Logit [69](#)  
  em modelos lineares generalizados mistos [50](#)  
ANOVA

ANOVA (*continuação*)  
  no GLM Multivariate [1](#)  
  no GLM Repeated Measures [9](#)  
ANOVA multivariado [1](#)

## B

Bonferroni  
  no GLM Multivariate [5](#)  
  no GLM Repeated Measures [14](#)

## C

C de Dunnett  
  no GLM Multivariate [5](#)  
  no GLM Repeated Measures [14](#)  
casos censurados  
  em Kaplan-Meier [92](#)  
  em Regressão de Cox [94](#)  
  em Tabelas de Mortalidade [72](#)  
categoria de referência  
  em Equações de Estimativa Generalizada [43](#), [44](#)  
  em Modelos Lineares Generalizados [32](#)  
classe geradora  
  em Análise Log-Linear de Seleção de Modelo [66](#)  
Componentes de variância  
  modelo [19](#)  
  opções [20](#)  
  recursos adicionais do comando [21](#)  
  salvando resultados [21](#)  
construir termos [3](#), [12](#), [19](#), [66](#), [68](#), [71](#)  
contrastes  
  em Análise Log-Linear Geral [67](#)  
  em Análise Log-Linear Logit [69](#)  
  em Regressão de Cox [95](#)  
convergência do log da verossimilhança  
  em Equações de Estimativa Generalizada [45](#)  
  em Modelos Lineares Generalizados [34](#)  
  em Modelos Lineares Mistos [26](#)  
Convergência hessiana  
  em Equações de Estimativa Generalizada [45](#)  
  em Modelos Lineares Generalizados [34](#)  
convergência paramétrica  
  em Equações de Estimativa Generalizada [45](#)  
  em Modelos Lineares Generalizados [34](#)  
  em Modelos Lineares Mistos [26](#)  
covariáveis  
  em Regressão de Cox [95](#)  
covariáveis de sequência de caracteres  
  em Regressão de Cox [95](#)  
covariáveis dependentes de tempo segmentadas  
  em Regressão de Cox [98](#)

## D

decomposição hierárquica

- decomposição hierárquica (*continuação*)
  - em Componentes de Variância [20](#)
- diferença menos significativa
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- diferença significativa honesta de Tukey
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- Distância de Cook
  - em Modelos Lineares Generalizados [37](#)
  - no GLM [6](#)
  - no GLM Repeated Measures [16](#)
- distribuição binomial
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- distribuição binomial negativa negativa
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- Distribuição de Poisson
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- Distribuição de Tweedie
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- distribuição gama
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- Distribuição Gaussiana inversa
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- distribuição multinomial
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- distribuição normal
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)

## E

- efeitos aleatórios
  - em Modelos Lineares Mistos [25, 28](#)
- efeitos fixos
  - em Modelos Lineares Mistos [24](#)
- eliminação backward
  - em Análise Log-Linear de Seleção de Modelo [65](#)
- Equações de estimativa generalizada
  - categoria de referência para resposta binária [43](#)
  - critérios de estimação [45](#)
  - especificação de modelo [44](#)
  - estatísticas [47](#)
  - exportação de modelo [49](#)
  - médias marginais estimadas [48](#)
  - opções para fatores categóricos [44](#)
  - preditores [44](#)
  - resposta [43](#)
  - salvar variáveis no conjunto de dados ativo [49](#)
  - tipo de modelo [41](#)
  - valores iniciais [46](#)
- erro padrão
  - no GLM [6](#)
  - no GLM Repeated Measures [16](#)
- escoragem
  - em Modelos Lineares Mistos [26](#)
- Escoragem de Fisher

- Escoragem de Fisher (*continuação*)
  - em Modelos Lineares Mistos [26](#)
- estatística Wald
  - em Análise Log-Linear Geral [67](#)
  - em Análise Log-Linear Logit [69](#)
- estatísticas descritivas
  - em Equações de Estimativa Generalizada [47](#)
  - em Modelos Lineares Generalizados [35](#)
  - em Modelos Lineares Mistos [27](#)
- estimação de máxima verossimilhança
  - em Componentes de Variância [20](#)
- estimação de máxima verossimilhança restrita
  - em Componentes de Variância [20](#)
- estimativas de parâmetro
  - em Análise Log-Linear de Seleção de Modelo [66](#)
  - em Análise Log-Linear Geral [67](#)
  - em Análise Log-Linear Logit [69](#)
  - em Equações de Estimativa Generalizada [47](#)
  - em Modelos Lineares Generalizados [35](#)
  - em Modelos Lineares Mistos [27](#)
- estruturas de covariâncias
  - em Modelos Lineares Mistos [102](#)
- etapa pela metade
  - em Equações de Estimativa Generalizada [45](#)
  - em Modelos Lineares Generalizados [34](#)
  - em Modelos Lineares Mistos [26](#)

## F

- F múltiplo de Ryan-Einot-Gabriel-Welsch
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- fatores
  - no GLM Repeated Measures [11](#)
- frequências
  - em Análise Log-Linear de Seleção de Modelo [66](#)
- função de ligação
  - modelos lineares generalizados mistos [53](#)
- função de ligação binomial negativa
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação de Cauchit acumulativa
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação de complemento de log
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação de identidade
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação de log
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação de poder
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação de poder de chances
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação de probito
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação de probito acumulativa
  - em equações de estimativa generalizada [41](#)



- função de ligação de probito acumulativa (*continuação*)
  - em modelos lineares generalizados [29](#)
- função de ligação log-log
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação log-log acumulativa
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação log-log negativa
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação log-log negativa acumulativa
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação logit
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de ligação logit acumulativa
  - em equações de estimativa generalizada [41](#)
  - em modelos lineares generalizados [29](#)
- função de sobrevivência
  - em Tabelas de Mortalidade [72](#)
- função estimável geral
  - em Equações de Estimativa Generalizada [47](#)
- funções estimáveis gerais
  - em Modelos Lineares Generalizados [35](#)

## G

- GLM
  - salvando matrizes [6](#)
  - salvando variáveis [6](#)
- GLM multivariado [1](#)
- GLM Multivariate
  - covariáveis [1](#)
  - fatores [1](#)
  - gráficos de perfil [5](#)
  - lista de variáveis dependentes, [1](#)
  - testes post hoc [5](#)
- GLM Repeated Measures
  - definir fatores [11](#)
  - gráficos de perfil [14](#)
  - modelo [12](#)
  - recursos adicionais do comando [17](#)
  - salvando variáveis [16](#)
  - testes post hoc [14](#)
- GLOR
  - em Análise Log-Linear Geral [67](#)
- gráficos
  - em Análise Log-Linear Geral [68](#)
  - em Análise Log-Linear Logit [71](#)
- gráficos de perfil
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- gráficos de probabilidade normal
  - em Análise Log-Linear de Seleção de Modelo [66](#)
- GT2 de Hochberg
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)

## H

- histórico de iteração

- histórico de iteração (*continuação*)
  - em Equações de Estimativa Generalizada [47](#)
  - em Modelos Lineares Generalizados [35](#)
  - em Modelos Lineares Mistos [26](#)

## I

- informações a priori de efeito aleatório
  - em Componentes de Variância [20](#)
- informações de modelo
  - em Equações de Estimativa Generalizada [47](#)
  - em Modelos Lineares Generalizados [35](#)
- informações de nível do fator
  - em Modelos Lineares Mistos [27](#)
- intervalos de confiança
  - em Análise Log-Linear Geral [68](#)
  - em Análise Log-Linear Logit [71](#)
  - em Modelos Lineares Mistos [27](#)
- iterações
  - em Análise Log-Linear de Seleção de Modelo [66](#)
  - em Equações de Estimativa Generalizada [45](#)
  - em Modelos Lineares Generalizados [34](#)

## K

- Kaplan-Meier
  - comparar níveis de fator [93](#)
  - definindo eventos [93](#)
  - estatísticas [92](#), [94](#)
  - exemplo [92](#)
  - gráficos [94](#)
  - quartis [94](#)
  - recursos adicionais do comando [94](#)
  - salvando novas variáveis [93](#)
  - tabelas de sobrevivência [94](#)
  - tempo de sobrevivência média e mediana [94](#)
  - tendência linear para níveis do fator [93](#)
  - variáveis de status de sobrevivência [93](#)

## L

- LSD de Fisher
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)

## M

- matriz de correlações
  - em Equações de Estimativa Generalizada [47](#)
  - em Modelos Lineares Generalizados [35](#)
  - em Modelos Lineares Mistos [27](#)
- matriz de covariância paramétrica
  - em Modelos Lineares Mistos [27](#)
- matriz de covariância residual
  - em Modelos Lineares Mistos [27](#)
- matriz de covariâncias
  - em Equações de Estimativa Generalizada [45](#), [47](#)
  - em Modelos Lineares Generalizados [34](#), [35](#)
  - em Modelos Lineares Mistos [27](#)
  - no GLM [6](#)
- matriz de covariâncias de efeito aleatório
  - em Modelos Lineares Mistos [27](#)
- matriz dos coeficientes de contraste

- matriz dos coeficientes de contraste (*continuação*)
  - em Equações de Estimativa Generalizada [47](#)
  - em Modelos Lineares Generalizados [35](#)
- matriz L.
  - em Equações de Estimativa Generalizada [47](#)
  - em Modelos Lineares Generalizados [35](#)
- médias marginais estimadas
  - em Equações de Estimativa Generalizada [48](#)
  - em Modelos Lineares Generalizados [36](#)
  - em Modelos Lineares Mistos [28](#)
- método Newton-Raphson
  - em Análise Log-Linear Geral [67](#)
  - em Análise Log-Linear Logit [69](#)
- MINQUE
  - em Componentes de Variância [20](#)
- modelo de risco proporcional
  - em Regressão de Cox [94](#)
- modelo linear generalizado
  - em modelos lineares generalizados mistos [50](#)
- modelo linear geral
  - modelos lineares generalizados mistos [50](#)
- modelos customizados
  - em Análise Log-Linear de Seleção de Modelo [66](#)
  - em Componentes de Variância [19](#)
  - no GLM Repeated Measures [12](#)
- Modelos de fragilidade compartilhados paramétricos
  - análise [81](#)
  - criar gráfico [86](#)
  - critérios [82](#)
  - estimativa [83](#)
  - exportar [87](#)
  - imprimir [85](#)
  - modelo [83](#)
  - Prever [85](#)
- Modelos De Tempo De Falha Acelerada Paramétrica
  - análise [74](#)
  - criar gráfico [79](#)
  - critérios [75](#)
  - estimativa [76](#)
  - exportar [80](#)
  - imprimir [78](#)
  - modelo [76](#)
  - Prever [78](#)
- modelos de vários níveis
  - modelos lineares generalizados mistos [50](#)
- modelos fatoriais completos
  - em Componentes de Variância [19](#)
  - no GLM Repeated Measures [12](#)
- modelos hierárquicos
  - modelos lineares generalizados mistos [50](#)
- Modelos lineares generalizados
  - categoria de referência para resposta binária [32](#)
  - critérios de estimação [34](#)
  - distribuição [29](#)
  - especificação de modelo [33](#)
  - estatísticas [35](#)
  - exportação de modelo [38](#)
  - função de ligação [29](#)
  - médias marginais estimadas [36](#)
  - opções para fatores categóricos [33](#)
  - preditores [32](#)
  - resposta [31](#)
  - salvar variáveis no conjunto de dados ativo [37](#)
  - tipos de modelo [29](#)

- Modelos lineares generalizados (*continuação*)
  - valores iniciais [35](#)
- modelos lineares generalizados mistos
  - bloco de efeito aleatório [56](#)
  - coeficientes fixos [62](#)
  - covariâncias de efeito aleatório [63](#)
  - distribuição de resposta [53](#)
  - efeitos aleatórios [56](#)
  - efeitos fixos [55](#), [62](#)
  - estrutura de dados [61](#)
  - exportação de modelo [60](#)
  - função de ligação [53](#)
  - médias estimadas [64](#)
  - médias marginais estimadas [59](#)
  - offset [57](#)
  - parâmetros de covariância [63](#)
  - ponderação de análise [57](#)
  - predito por observado [61](#)
  - salvar campos [60](#)
  - sumarização do modelo [61](#)
  - tabela de classificação [62](#)
  - termos customizados [56](#)
  - visualização do modelo [61](#)
- modelos log-lineares hierárquicos [65](#)
- modelos logit de multinômio [69](#)
- modelos longitudinais
  - modelos lineares generalizados mistos [50](#)
- modelos mistos
  - linear [21](#)
  - modelos lineares generalizados mistos [50](#)
- Modelos mistos lineares
  - construir termos [24](#)
  - critérios de estimação [26](#)
  - efeitos aleatórios [25](#), [28](#)
  - efeitos fixos [24](#)
  - estruturas de covariâncias [102](#)
  - médias marginais estimadas [28](#)
  - modelo [27](#)
  - recursos adicionais do comando [28](#)
  - salvar variáveis [28](#)
  - termos de interação [24](#)
- Modelos paramétricos de fragilidade
  - variáveis de status de sobrevivência [87](#)
- modelos saturados
  - em Análise Log-Linear de Seleção de Modelo [66](#)

## N

- Newman-Keuls
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)

## P

- parâmetro de escala
  - em Equações de Estimativa Generalizada [45](#)
  - em Modelos Lineares Generalizados [34](#)
- proporção fracionária
  - em Análise Log-Linear Geral [67](#)

## Q

- qualidade do ajuste

qualidade do ajuste (*continuação*)  
em Equações de Estimativa Generalizada [47](#)  
em Modelos Lineares Generalizados [35](#)

## R

R-E-G-W F  
no GLM Multivariate [5](#)  
no GLM Repeated Measures [14](#)  
R-E-G-W Q  
no GLM Multivariate [5](#)  
no GLM Repeated Measures [14](#)  
razão de possibilidades de log generalizada  
em Análise Log-Linear Geral [67](#)  
Regressão de Cox  
contrastes [95](#)  
covariáveis [94](#)  
covariáveis categóricas [95](#)  
covariáveis de sequência de caracteres [95](#)  
covariáveis dependentes de tempo [98](#)  
definir evento [97](#)  
DfBeta(s) [96](#)  
entrada e remoção stepwise [97](#)  
estatísticas [94](#), [97](#)  
exemplo [94](#)  
função de risco [96](#)  
função de sobrevivência [96](#)  
funções da linha de base [97](#)  
gráficos [96](#)  
iterações [97](#)  
recursos adicionais do comando [97](#)  
resíduo parcial [96](#)  
salvando novas variáveis [96](#)  
variável de status de sobrevivência [97](#)  
Regressão de Kernel Ridge  
dobras de validação cruzada [138](#)  
exibição [138](#)  
gráficos [138](#)  
Parâmetros [138](#)  
parâmetros da grade [138](#)  
salvar [138](#)  
regressão de Poisson  
em Análise Log-Linear Geral [67](#)  
modelos lineares generalizados mistos [50](#)  
regressão logística  
modelos lineares generalizados mistos [50](#)  
regressão logística multinomial  
modelos lineares generalizados mistos [50](#)  
regressão multivariada [1](#)  
resíduos  
em Análise Log-Linear de Seleção de Modelo [66](#)  
em Análise Log-Linear Geral [69](#)  
em Análise Log-Linear Logit [72](#)  
em Equações de Estimativa Generalizada [49](#)  
em Modelos Lineares Generalizados [37](#)  
em Modelos Lineares Mistos [28](#)  
resíduos de deviance  
em Modelos Lineares Generalizados [37](#)  
resíduos de Pearson  
em Equações de Estimativa Generalizada [49](#)  
em Modelos Lineares Generalizados [37](#)  
resíduos de probabilidade  
em Modelos Lineares Generalizados [37](#)  
resíduos excluídos

resíduos excluídos (*continuação*)  
no GLM [6](#)  
no GLM Repeated Measures [16](#)  
resíduos não padronizados  
no GLM [6](#)  
no GLM Repeated Measures [16](#)  
resíduos padronizados  
no GLM [6](#)  
no GLM Repeated Measures [16](#)  
Ridge do kernel  
alfa [136](#)  
coef0 [136](#)  
gama [136](#)  
grau [136](#)  
modelo único [136](#)  
seleção de modelo [136](#)

## S

separação  
em Equações de Estimativa Generalizada [45](#)  
em Modelos Lineares Generalizados [34](#)  
Sobrevivência AFT  
Diálogo sobrevivência-Variáveis de categoria  
[80](#)  
soma dos quadrados [3](#), [12](#)  
somadas de quadrados  
em Componentes de Variância [20](#)  
em Modelos Lineares Mistos [25](#)  
Student-Newman-Keuls  
no GLM Multivariate [5](#)  
no GLM Repeated Measures [14](#)  
sumarização de processamento de caso  
em Equações de Estimativa Generalizada [47](#)  
em Modelos Lineares Generalizados [35](#)

## T

T2 de Tamhane  
no GLM Multivariate [5](#)  
no GLM Repeated Measures [14](#)  
T3 de Dunnett  
no GLM Multivariate [5](#)  
no GLM Repeated Measures [14](#)  
tabelas de contingência  
em Análise Log-Linear Geral [67](#)  
Tabelas de mortalidade  
comparar níveis de fator [74](#)  
estatísticas [72](#)  
exemplo [72](#)  
função de sobrevivência [72](#)  
gráficos [74](#)  
recursos adicionais do comando [74](#)  
suprimindo exibição da tabela [74](#)  
taxa de risco [72](#)  
teste de Wilcoxon (Gehan) [74](#)  
variáveis de fator [74](#)  
variáveis de status de sobrevivência [80](#)  
tabulação cruzada  
em Análise Log-Linear de Seleção de Modelo [65](#)  
taxa de risco  
em Tabelas de Mortalidade [72](#)  
termos aninhados

- termos aninhados (*continuação*)
  - em Equações de Estimativa Generalizada [44](#)
  - em Modelos Lineares Generalizados [33](#)
  - em Modelos Lineares Mistos [24](#)
- termos de interação
  - em Modelos Lineares Mistos [24](#)
- teste b de Tukey
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- teste de amplitude múltipla de Duncan
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- teste de Breslow
  - em Kaplan-Meier [93](#)
- teste de comparação entre pares de Gabriel
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- teste de comparações pairwise de Games-Howell
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- teste de Gehan
  - em Tabelas de Mortalidade [74](#)
- teste de ranqueamento de log
  - em Kaplan-Meier [93](#)
- teste de Scheffé
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- teste de Tarone-Ware
  - em Kaplan-Meier [93](#)
- teste de Wilcoxon
  - em Tabelas de Mortalidade [74](#)
- teste dos multiplicadores de Lagrange
  - em Modelos Lineares Generalizados [35](#)
- teste paramétrica de covariância
  - em Modelos Lineares Mistos [27](#)
- teste t de Dunnett
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- teste t de Sidak
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- teste t de Waller-Duncan
  - no GLM Multivariate [5](#)
  - no GLM Repeated Measures [14](#)
- tolerância a singularidade
  - em Modelos Lineares Mistos [26](#)

## V

- valores de ponto de alavanca
  - em Modelos Lineares Generalizados [37](#)
  - no GLM [6](#)
  - no GLM Repeated Measures [16](#)
- valores preditos
  - em Análise Log-Linear Geral [69](#)
  - em Análise Log-Linear Logit [72](#)
  - em Modelos Lineares Mistos [28](#)
- valores preditos fixos
  - em Modelos Lineares Mistos [28](#)
- valores preditos ponderados
  - no GLM [6](#)
  - no GLM Repeated Measures [16](#)
- variáveis de assunto
  - em Modelos Lineares Mistos [22](#)

- variáveis de medidas repetidas
  - em Modelos Lineares Mistos [22](#)
- visualização do modelo
  - em modelos lineares generalizados mistos [61](#)



