

IBM SPSS Advanced Statistics 24

IBM

Comunicado

Antes de usar estas informações e o produto suportado por elas, leia as informações nos “Avisos” na página 105.

Informações sobre o produto

Esta edição aplica-se à versão 24, liberação 0, modificação 0 do IBM SPSS Statistics e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

Índice

Capítulo 1. Introdução às Estatísticas Avançadas 1

Capítulo 2. Análise do GLM Multivariado 3

Modelo GLM Multivariate	4
Termos de compilação	4
Soma dos Quadrados	5
Contrastes do GLM Multivariate	6
Tipos de Contraste	6
Gráficos de Perfil do GLM Multivariate	6
Comparações Post Hoc do GLM Multivariate	7
Salvamento de GLM	8
Opções do GLM Multivariado	9
Recursos Adicionais de Comando do GLM	10

Capítulo 3. GLM Repeated Measures . . 11

Medidas Repetidas do GLM - Definir Fatores	13
Modelo de Medidas Repetidas do GLM	14
Termos de compilação	14
Soma dos Quadrados	14
Contrastes do GLM Repeated Measures	15
Tipos de Contraste	15
Contrastes de Perfil do GLM Repeated Measures	16
Comparações Post Hoc do GLM Repeated Measures	16
Salvamento de Medidas Repetidas do GLM	18
Opções de Medidas Repetidas do GLM	18
Recursos Adicionais de Comando do GLM	19

Capítulo 4. Análise de Componentes de Variância 21

Modelos de Componentes de Variância	22
Termos de compilação	22
Opções de Componentes de Variância	22
Soma dos Quadrados (Componentes de Variância)	23
Componentes de Variância Salvos no Novo Arquivo	24
Recursos Adicionais do Comando VARCOMP	24

Capítulo 5. Modelos mistos lineares . . 25

Modelos Lineares Mistos - Selecionar Assuntos/Variáveis Repetidas	26
Efeitos Fixos de Modelos Lineares Mistos	27
Construir Termos Não Aninhados	27
Construir Termos Aninhados	27
Soma dos Quadrados	27
Efeitos Aleatórios de Modelos Lineares Mistos	28
Estimação de Modelos Lineares	29
Estatísticas de Modelos Lineares Mistos	29
Médias de EM de Modelos Lineares Mistos	30
Salvamento de Modelos Lineares Mistos	30
Recursos Adicionais do Comando MIXED	31

Capítulo 6. Modelos lineares generalizados 33

Resposta de Modelos Lineares Generalizados	35
Categoria de Referência de Modelos Lineares Generalizados	36
Preditores de Modelos Lineares Generalizados	36
Opções de Modelos Lineares Generalizados	37
Modelo de Modelos Lineares Generalizados	37
Estimação de Modelos Lineares Generalizados	38
Valores de Modelos Lineares Generalizados	39
Estatísticas de Modelos Lineares Generalizados	39
Médias de EM de Modelos Lineares Generalizados	41
Salvamento de Modelos Lineares Generalizados	42
Exportação de Modelos Lineares Generalizados	43
Variáveis Adicionais de Comando do GENLIN	43

Capítulo 7. Equações de Estimativa Generalizada 45

Tipo de Modelo de Equações de Estimativa Generalizada	47
Resposta de Equações de Estimativa Generalizada	49
Categoria de Referência de Equações de Estimativa Generalizada	49
Preditores de Equações de Estimativa Generalizada	49
Opções de Equações de Estimativa Generalizada	50
Modelo de Equações de Estimativa Generalizada	50
Estimação de Equações de Estimativa Generalizada	51
Valores Iniciais de Equações de Estimativa Generalizada	52
Estatísticas de Equações de Estimativa Generalizada	53
Médias de EM de Equações de Estimativa Generalizada	54
Salvamento de Equações de Estimativa Generalizada	55
Exportação de Equações de Estimativa Generalizada	55
Variáveis Adicionais de Comando do GENLIN	56

Capítulo 8. Modelos lineares generalizados mistos 57

Obtendo um modelo linear generalizado misto	58
Resposta	58
Efeitos Fixos	60
Incluir um Termo Customizado	61
Efeitos Aleatórios	62
Bloco de Efeito Aleatório	62
Ponderação e Offset	63
Opções de Criação Geral	63
Estimação	64
Médias Estimadas	65
Salvar	66
Visualização do modelo	66
Sumarização do Modelo	67
Estrutura de Dados	67
Predito por Observado	67
Classificação	67

Efeitos Fixos	68	Kaplan-Meier - Comparar Níveis do Fator	88
Coeficientes Fixos	68	Kaplan-Meier - Salvar Novas Variáveis	88
Covariâncias de Efeito Aleatório	69	Opções de Kaplan-Meier	89
Parâmetros de Covariância	69	Variáveis Adicionais de Comando do KM	89
Médias Estimadas: Efeitos Significativos	69		
Médias Estimadas: Efeitos Customizados	70		
Capítulo 9. Análise de log linear de seleção do modelo	71	Capítulo 14. Análise de Regressão de Cox	91
Análise Log-Linear - Definir Intervalo	72	Regressão de Cox - Definir Variáveis Categóricas	92
Modelo de Análise Log-Linear	72	Gráficos de Regressão de Cox	92
Termos de compilação	72	Regressão de Cox - Salvar Novas Variáveis	93
Opções de Análise Log-Linear de Seleção de Modelo	72	Opções de Regressão de Cox	93
Recursos Adicionais do Comando HILOGLINEAR	73	Regressão de Cox - Definir Evento para Variável de Status	94
		Variáveis Adicionais de Comando do COXREG	94
Capítulo 10. Análise de log linear geral	75	Capítulo 15. Calculando Covariáveis Dependentes de Tempo	95
Modelo de Análise Log-Linear Geral	76	Calculando uma Covariável Dependente de Tempo	95
Termos de compilação	76	Regressão de Cox com Variáveis Adicionais de Covariáveis Dependentes de Tempo	96
Opções de Análise Log-Linear Geral	76		
Salvamento de Análise Log-Linear Geral	77	Capítulo 16. Esquemas de codificação de variável categórica	97
Recursos Adicionais do Comando GENLOG	77	Desvio	97
		Simplex	97
Capítulo 11. Análise Log-Linear Logit	79	Helmert	98
Modelo de Análise Log-Linear Logit	80	Diferença	98
Termos de compilação	80	Polinomial	98
Opções de Análise Log-Linear Logit	81	Repetido	99
Salvamento de Análise Log-Linear Logit	81	Especial	99
Recursos Adicionais do Comando GENLOG	82	Indicador	100
		Capítulo 17. Estruturas de Covariâncias.	101
Capítulo 12. Tabelas de mortalidade	83	Avisos	105
Definir Eventos de Tabelas de Mortalidade para Variáveis de Status	84	Marcas comerciais	107
Definir Intervalo para Tabelas de Mortalidade	84		
Opções de Tabelas de Mortalidade	84	Índice Remissivo	109
Variáveis Adicionais de Comando do SURVIVAL	85		
Capítulo 13. Análise de Sobrevida de Kaplan-Meier.	87		
Kaplan-Meier - Definir Evento para Variável de Status	88		

Capítulo 1. Introdução às Estatísticas Avançadas

A opção Estatísticas Avançadas fornece procedimentos que oferecem opções de modelagem mais avançadas que estão disponíveis por meio da opção Estatísticas de Base.

- 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.

Capítulo 2. 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 (preditoras) são especificadas como covariáveis.

Modelos balanceados e não balanceados podem ser testados. Um design será balanceado se cada célula no modelo contiver o mesmo número de casos. 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 resíduos 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 efeitos principais do fator, todos os efeitos principais de covariáveis e todas as interações fator por fator. Ele não contém interações de covariáveis. Selecione **Customizado** para especificar apenas um subconjunto de interações ou para especificar interações fator por covariável. Deve-se indicar todos os termos a serem incluídos no modelo.

Fatores e Covariáveis. Os fatores e covariáveis são listados.

Modelo. O modelo depende da natureza de seus dados. 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 nenhuma célula omissa, o método da soma dos quadrados do Tipo III é o mais normalmente utilizado.

Incluir intercepto no modelo. O intercepto geralmente é incluído no modelo. Se você conseguir presumir que os dados passam pela origem, será possível excluir o intercepto.

Termos de compilação

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

Todas 3 fatores. Cria todas as interações de três fatores 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.

Soma dos Quadrados

Para o modelo, é possível escolher um tipo de somas de quadrados. O tipo III é o mais comumente usado e é o padrão.

Tipo I. Esse método também é conhecido como decomposição hierárquica do método da soma dos quadrados. Cada termo é ajustado para apenas o termo que o precede no modelo. As somas dos quadrados do tipo I são comumente utilizados 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 no qual o efeito primeiro especificado é aninhado dentro do segundo efeito especificado, o efeito segundo especificado é aninhado dentro do terceiro, e assim por diante. (Esta forma de aninhamento pode ser especificada usando 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. A soma dos quadrados do Tipo III possui uma vantagem principal nas quais eles são invariáveis com relação às frequências de célula contanto que o formato geral de estimabilidade permaneça constante. Portanto, esse tipo de somas de quadrados é geralmente considerado útil para um modelo não balanceado sem células omissas. Em um design fatorial sem células omissas, este método é equivalente à técnica Yates' weighted-squares-of-means. O método da soma dos quadrados do Tipo III é comumente utilizado para:

- Todos os modelos listados em Tipo I e Tipo II.
- Qualquer modelo balanceado ou 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 utilizados para estimar as 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 linhas em que cada ponto indica a média marginal estimada de uma variável dependente (ajustada para quaisquer covariáveis) em um nível de um fator. Os níveis de um segundo fator podem ser utilizados para tornar as linhas separadas. Cada nível em um terceiro fator pode ser utilizado 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 os níveis. Para dois ou mais fatores, linhas paralelas indicam que não existe nenhuma interação entre fatores, o que significa que é possível investigar os níveis somente de um fator. Linhas

não paralelas indicam uma interação.

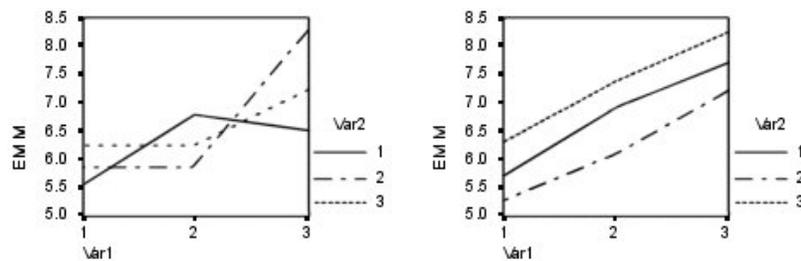


Figura 1. Gráfico não paralelo (esquerdo) e gráfico paralelo (direito)

Após um gráfico ser especificado ao selecionar os fatores para o eixo horizontal e, opcionalmente, os fatores para linhas e gráficos separados, o gráfico deverá 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 não ajustados. Os testes post hoc são executados para cada variável dependente separadamente.

Os testes da diferença significativa honesta de Bonferroni e Tukey são geralmente utilizados para testes de comparação múltipla. O **teste Bonferroni**, com base na estatística t de Student, ajusta o nível de significância observado para o fato de que diversas comparações são feitas. O **teste t de Sidak** também ajusta o nível de significância e fornece limites mais apertados do que o teste de Bonferroni. O **teste da diferença significativa honesta de Tukey** utiliza o intervalo estatístico estudentizado para fazer todas as comparações entre pares entre os grupos e configura a taxa de erro entre experimentos para a 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 utilizado. Geralmente, o teste de Tukey é mais poderoso. O **teste de comparação entre pares de Gabriel** também utiliza o módulo máximo estudentizado e é geralmente mais poderoso que o GT2 de Hochberg quando os tamanhos de células são desiguais. O teste de Gabriel pode ser liberal quando os tamanhos de células variam grandiosamente.

O **teste t de comparação múltipla entre pares de Dunnett** compara um conjunto de tratamentos com relação a uma média de controle única. A última categoria é a categoria de controle padrão. Como alternativa, é possível escolher a primeira categoria. Também é possível escolher um teste bilateral ou unilateral. Para comprovar que a média em qualquer nível (exceto a categoria de controle) do fator não é igual à da categoria de controle, utilize um teste bilateral. Para testar se a média em qualquer nível do fator é menor do que a da categoria de controle, selecione **< Controle**. Da mesma forma, para testar se a média em qualquer nível do fator é maior do que a da categoria de controle, selecione **> Controle**.

Ryan, Einot, Gabriel e Welsch (R-E-G-W) desenvolveram dois testes de intervalo step-down múltiplos. Os procedimentos step-down múltiplos primeiro testam se todas as médias são iguais. Se todas as médias não forem iguais, os subconjuntos de médias serão testados para igualdade. **R-E-G-W F** baseia-se em um teste F e **R-E-G-W Q** baseia-se 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 step-down), mas não são recomendados para tamanhos de células desiguais.

Quando as variâncias são desiguais, utilize **T2 de Tamhane** (testes de comparações entre pares conservadores com base em um teste t), **T3 de Dunnett** (teste de comparação entre pares com base no

módulo máximo estudentizado), **Games-Howell Teste de comparação entre pares** (às vezes liberal) ou **C de Dunnett** (teste de comparação entre pares com base no intervalo estudentizado).

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

O **teste t de Waller-Duncan** utiliza abordagem Bayesiana. Esse teste de intervalo utiliza 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 sejam testadas, não apenas as comparações entre pares disponíveis neste recurso. O resultado é que o teste de Scheffé é geralmente mais conservador do que outros testes, significando que uma diferença maior entre as médias é necessária para significância.

O teste de comparação múltipla entre pares da diferença menos significativa (**LSD**) é equivalente a diversos testes *t* individuais entre todos os pares de grupos. A desvantagem deste teste é que nenhuma tentativa é feita para ajustar o nível de significância observado para comparações múltiplas.

Testes exibidos. Comparações entre pares são fornecidas para LSD, Sidak, Bonferroni, Games-Howell, T2 e T3 de Tamhane, C de Dunnett e T3 de Dunnett. Subconjuntos homogêneos para testes de intervalo são fornecidos para S-N-K, *b* de Tukey, Duncan, *F* de R-E-G-W, *Q* de R-E-G-W e Waller. O teste da diferença significativa honesta de Tukey, GT2 de Hochberg, teste de Gabriel e teste de Scheffé são tanto testes de comparação múltipla quanto testes de intervalo.

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 Statistics, deve-se salvar o arquivo de dados atual

Valores Preditos. Os valores que o modelo prediz para cada caso.

- *Não padronizado.* O valor que o modelo prediz para a variável dependente.
- *Ponderado.* Valores preditos não padronizados ponderados. Disponível apenas se uma variável WLS foi selecionada anteriormente.
- *Erro padrão.* Uma estimativa do desvio padrão do valor médio da variável dependente para os casos que tiverem os mesmos valores das variáveis independentes.

Diagnósticos. Medidas para identificar os casos com combinações de valores incomuns para as variáveis independentes e os casos que podem ter um grande impacto no modelo.

- *Distância de cook.* Uma medida do quanto os resíduos de todos os casos seriam alterados se um caso específico 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.
- *REMOVE.* Valores de ponto de alavanca não centralizados. A influência relativa de cada observação sobre o ajuste do modelo.

Residuais. Um resíduo não padronizado é o valor real da variável dependente menos o valor predito pelo modelo. Resíduos padronizados, estudentizados e excluídos também estão disponíveis. Se uma variável WLS foi escolhida, 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 do seu desvio padrão. Resíduos padronizados, também conhecidos como resíduos de Pearson, possuem uma média de 0 e um desvio padrão de 1.
- *Estudentizado*. O resíduo dividido por uma estimativa do seu desvio padrão que varia de caso para caso, dependendo da distância dos valores de cada caso nas variáveis independentes das médias das variáveis independentes.
- *Excluído*. O resíduo para um caso quando esse caso é excluído do cálculo dos coeficientes de regressão. Ele é a diferença entre o valor da variável dependente e o valor predito ajustado.

Estatísticas de coeficiente. Grava uma matriz de variância-covariância das estimativas do parâmetro no modelo em um novo conjunto de dados na sessão atual ou em um arquivo de dados externo do IBM SPSS Statistics. Além disso, para cada variável dependente, haverá uma linha de estimativas 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 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.

Médias Marginais Estimadas. 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. Interações estão disponíveis somente se você tiver especificado um modelo customizado.

- **Compare os 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.
- **Ajustamento de intervalo de confiança.** Selecione o ajustamento de diferença menos significativa (LSD), Bonferroni ou Sidak ou para os intervalos de confiança e significância. Este item estará disponível apenas se **Comparar os efeitos principais** for selecionada.

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 η^2 quadrado parcial para cada efeito e cada estimativa paramétrica. A estatística η^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 paramétrica** para produzir as estimativas paramétrica, 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. A **função estimável geral** permite construir testes de hipótese customizados com base na função estimável geral. As linhas em qualquer matriz de coeficientes de contraste são combinações lineares da função estimável geral.

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 na caixa de 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.

Capítulo 3. 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) dentro-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 e comparações múltiplas a posteriori (para fatores entre assuntos): 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. 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 Caixa; 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 The 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 *ponderação1*, ..., *ponderação5* são usadas 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 efeitos principais do fator, todos os efeitos principais de covariáveis e todas as interações fator por fator. Ele não contém interações de covariáveis. Selecione **Customizado** para especificar apenas um subconjunto de interações ou para especificar interações fator por covariável. Deve-se indicar todos os termos a serem incluídos no modelo.

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 compilação

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

Todas 3 fatores. Cria todas as interações de três fatores 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.

Soma dos Quadrados

Para o modelo, é possível escolher um tipo de somas de quadrados. O tipo III é o mais comumente usado e é o padrão.

Tipo I. Esse método também é conhecido como decomposição hierárquica do método da soma dos quadrados. Cada termo é ajustado para apenas o termo que o precede no modelo. As somas dos quadrados do tipo I são comumente utilizados 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 no qual o efeito primeiro especificado é aninhado dentro do segundo efeito especificado, o efeito segundo especificado é aninhado dentro do terceiro, e assim por diante. (Esta forma de aninhamento pode ser especificada usando 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. A soma dos quadrados do Tipo III possui uma vantagem principal nas quais eles são invariáveis com relação às frequências de célula contanto que o formato geral de estimabilidade permaneça constante. Portanto, esse tipo de somas de quadrados é geralmente considerado útil para um modelo não balanceado sem células omissas. Em um design fatorial sem células omissas, este método é equivalente à técnica Yates' weighted-squares-of-means. O método da soma dos quadrados do Tipo III é comumente utilizado para:

- Todos os modelos listados em Tipo I e Tipo II.
- Qualquer modelo balanceado ou 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 utilizados para estimar as 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 linhas em que cada ponto indica a média marginal estimada de uma variável dependente (ajustada para quaisquer covariáveis) em um nível de um fator. Os níveis de um segundo fator podem ser utilizados para tornar as linhas separadas. Cada nível em um terceiro fator pode ser utilizado 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 os níveis. Para dois ou mais fatores, linhas paralelas indicam que não existe nenhuma interação entre fatores, o que significa que é possível investigar os níveis somente de um fator. Linhas não paralelas indicam uma interação.

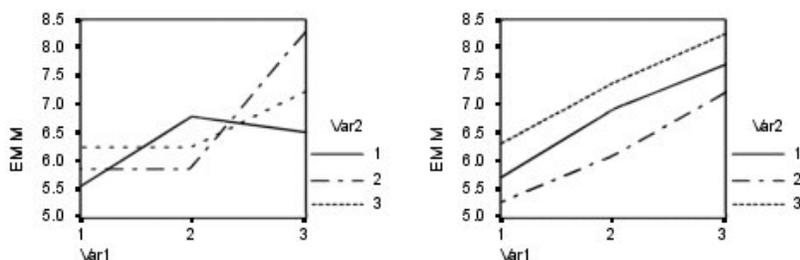


Figura 2. Gráfico não paralelo (esquerdo) e gráfico paralelo (direito)

Após um gráfico ser especificado ao selecionar os fatores para o eixo horizontal e, opcionalmente, os fatores para linhas e gráficos separados, o gráfico deverá 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 não ajustados. 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 são geralmente utilizados para testes de comparação múltipla. O **teste Bonferroni**, com base na estatística t de Student, ajusta o nível de significância observado para o fato de que diversas comparações são feitas. O **teste t de Sidak** também ajusta o nível de significância e fornece limites mais apertados do que o teste de Bonferroni. O **teste da diferença significativa honesta de Tukey** utiliza o intervalo estatístico estudentizado para fazer todas as comparações entre pares entre os grupos e configura a taxa de erro entre experimentos para a 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 utilizado. Geralmente, o teste de Tukey é mais poderoso. O **teste de comparação entre pares de Gabriel** também utiliza o módulo máximo estudentizado e é geralmente mais poderoso que o GT2 de Hochberg quando os tamanhos de células são desiguais. O teste de Gabriel pode ser liberal quando os tamanhos de células variam grandiosamente.

O **teste t de comparação múltipla entre pares de Dunnett** compara um conjunto de tratamentos com relação a uma média de controle única. A última categoria é a categoria de controle padrão. Como alternativa, é possível escolher a primeira categoria. Também é possível escolher um teste bilateral ou unilateral. Para comprovar que a média em qualquer nível (exceto a categoria de controle) do fator não é igual à da categoria de controle, utilize um teste bilateral. Para testar se a média em qualquer nível do fator é menor do que a da categoria de controle, selecione **< Controle**. Da mesma forma, para testar se a média em qualquer nível do fator é maior do que a da categoria de controle, selecione **> Controle**.

Ryan, Einot, Gabriel e Welsch (R-E-G-W) desenvolveram dois testes de intervalo step-down múltiplos. Os procedimentos step-down múltiplos primeiro testam se todas as médias são iguais. Se todas as médias não forem iguais, os subconjuntos de médias serão testados para igualdade. **R-E-G-W F** baseia-se em um teste *F* e **R-E-G-W Q** baseia-se 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 step-down), mas não são recomendados para tamanhos de células desiguais.

Quando as variâncias são desiguais, utilize **T2 de Tamhane** (testes de comparações entre pares conservadores com base em um teste *t*), **T3 de Dunnett** (teste de comparação entre pares com base no módulo máximo estudentizado), **Games-Howell Teste de comparação entre pares** (às vezes liberal) ou **C de Dunnett** (teste de comparação entre pares com base no intervalo estudentizado).

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

O **teste t de Waller-Duncan** utiliza abordagem Bayesiana. Esse teste de intervalo utiliza 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 sejam testadas, não apenas as comparações entre pares disponíveis neste recurso. O resultado é que o teste de Scheffé é geralmente mais conservador do que outros testes, significando que uma diferença maior entre as médias é necessária para significância.

O teste de comparação múltipla entre pares da diferença menos significativa (**LSD**) é equivalente a diversos testes *t* individuais entre todos os pares de grupos. A desvantagem deste teste é que nenhuma tentativa é feita para ajustar o nível de significância observado para comparações múltiplas.

Testes exibidos. Comparações entre pares são fornecidas para LSD, Sidak, Bonferroni, Games-Howell, T2 e T3 de Tamhane, C de Dunnett e T3 de Dunnett. Subconjuntos homogêneos para testes de intervalo são fornecidos para S-N-K, *b* de Tukey, Duncan, *F* de R-E-G-W, *Q* de R-E-G-W e Waller. O teste da diferença significativa honesta de Tukey, GT2 de Hochberg, teste de Gabriel e teste de Scheffé são tanto testes de comparação múltipla quanto testes de intervalo.

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 Statistics, deve-se salvar o arquivo de dados atual

Valores Preditos. Os valores que o modelo prediz para cada caso.

- *Não Padronizados.* 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 do quanto os resíduos de todos os casos seriam alterados se um caso específico 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 ponto de alavanca 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 Padronizados.* A diferença entre um valor observado e o valor predito pelo modelo
- *Padronizado.* O resíduo dividido por uma estimativa do seu desvio padrão. Resíduos padronizados, também conhecidos como resíduos de Pearson, possuem uma média de 0 e um desvio padrão de 1.
- *Estudentizado.* O resíduo dividido por uma estimativa do seu desvio padrão que varia de caso para caso, dependendo da distância dos valores de cada caso nas variáveis independentes das médias das variáveis independentes.
- *Excluído.* O resíduo para um caso quando 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.

Médias Marginais Estimadas. 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. Ambos os fatores dentre-sujeitos e entre assuntos podem ser selecionados.

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

- **Ajustamento de intervalo de confiança.** Selecione o ajustamento de diferença menos significativa (LSD), Bonferroni ou Sidak ou para os intervalos de confiança e significância. Este item estará disponível apenas se **Comparar os efeitos principais** for selecionada.

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 quadrado parcial para cada efeito e cada estimativa paramétrica. A estatística eta quadrado descreve a proporção da variabilidade total atribuível a um fator. Selecione **Potência observada** para obter a potência do teste quando a hipótese alternativa é configurada com base no valor observado. Selecione **Estimativas paramétrica** para produzir as estimativas paramétrica, os erros padrão, testes t , intervalos de confiança e a potência observada para cada teste. 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. A **função estimável geral** permite construir testes de hipótese customizados com base na função estimável geral. As linhas em qualquer matriz de coeficientes de contraste são combinações lineares da função estimável geral.

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 na caixa de 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.

Capítulo 4. 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 efeitos principais do fator, todos os efeitos principais de covariáveis e todas as interações fator por fator. Ele não contém interações de covariáveis. Selecione **Customizado** para especificar apenas um subconjunto de interações ou para especificar interações fator por covariável. Deve-se indicar todos os termos a serem incluídos no modelo.

Fatores e Covariáveis. Os fatores e covariáveis são listados.

Modelo. O modelo depende da natureza de seus dados. 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.

Incluir intercepto no modelo. Normalmente o intercepto é incluído no modelo. Se você conseguir presumir que os dados passam pela origem, será possível excluir o intercepto.

Termos de compilação

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

Todas 3 fatores. Cria todas as interações de três fatores 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.

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 equilibrado, em que quaisquer efeitos principais são 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 em que 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 Statistics.

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.

Capítulo 5. 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 paramétrica 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 de 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 às 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 Combinados > Linear...

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

3. Opcionalmente, selecione uma ou mais variáveis repetidas.

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

5. Clique em **Continuar**.
6. Selecione uma variável dependente.
7. Selecione pelo menos um fator ou covariável.
8. 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 - Selecionar Assuntos/Variáveis Repetidas

Essa caixa de diálogo lhe permite selecionar variáveis que definem assuntos e observações repetidas e escolher uma estruturas de covariâncias para os resíduos.

Assuntos. 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 de Assuntos são usadas para definir assuntos para a estruturas de covariâncias de resíduo. 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)
- 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: Correlações

Consulte o tópico Capítulo 17, “Estruturas de Covariâncias”, na página 101 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 construir 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. Esse 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.

Todas 2 fatores. Cria todas as interações de dois fatores possíveis das variáveis selecionadas.

Todas 3 fatores. Cria todas as interações de três fatores das variáveis selecionadas.

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

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

Construir Termos Aninhados

É possível construir termos aninhados para o seu modelo nesse procedimento. Termos aninhados são úteis para modelar o efeito de um fator ou de uma covariável cujos valores não interagem com os níveis de um 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**.

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

Limitações. Termos aninhados têm 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. Assim, se A for um fator e X uma covariável, então especificar $A(X)$ será 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 equilibrado, em que quaisquer efeitos principais são 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 em que 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

Consulte o tópico Capítulo 17, “Estruturas de Covariâncias”, na página 101 para obter mais informações

Efeitos Aleatórios. Não há nenhum modelo padrão; portanto, deve-se especificar explicitamente os efeitos aleatórios. Como alternativa, é possível construir 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.

Estimação de Modelos Lineares

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

Iterações: As opções a seguir estão disponíveis:

- **Máximo de iterações.** Especifique um número inteiro não negativo.
- **Divisões máximas da etapa pela metade.** Em cada iteração, o tamanho do passo é reduzido por um fator de 0,5 até que o log da verossimilhança aumente ou a divisão máxima da etapa pela metade seja atingida. Especifique um número inteiro positivo.
- **Imprimir histórico de iteração para cada n passo(s).** Exibe uma tabela contendo o valor da função de log da verossimilhança e as estimativas paramétrica em cada n iteração, iniciando com a 0^{ésima} 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 a mudança relativa na função de log da verossimilhança for menor que o valor especificado, que deve ser não negativo. O critério não será utilizado se o valor especificado for igual a 0.

Convergência paramétrica. A convergência será assumida se a mudança máxima absoluta ou a mudança máxima relativa nas estimativas paramétrica for menor 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.

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

Tolerância à singularidade. Este valor é utilizado 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 do 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 paramétrica.** Exibe as estimativas paramétrica de efeitos fixos e de efeitos aleatórios 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 entre fatores que foram especificados na caixa de diálogo Fixo, mais um termo GERAL. 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 GERAL for selecionado, as médias marginais estimadas da variável dependente serão exibidas, caindo sobre 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 e 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.

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.

Capítulo 6. 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 predizer 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 seguintes opções 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 seguintes opções 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 seguintes opções 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 seguintes opções 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 seguintes opções 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.

Customizada. 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 omissos, 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 Φ^{-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.** $f(x)=[(x/(1-x))^\alpha - 1]/\alpha$, if $\alpha \neq 0$. $f(x)=\log(x)$, if $\alpha=0$. for a especificação de número necessária e deve ser um número real. Isso é apropriado apenas com a distribuição binomial.
- **Probit.** $f(x)=\Phi^{-1}(x)$, em que Φ^{-1} é a função de distribuição acumulativa normal padrão inversa. Isso é apropriado apenas com a distribuição binomial.
- **Poder.** $f(x)=x^\alpha$, if $\alpha \neq 0$. $f(x)=\log(x)$, if $\alpha=0$. for a especificação de número 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. Fatores devem ter valores válidos para um caso a ser incluído 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 construir 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.

Todas 2 fatores. Cria todas as interações de dois fatores possíveis das variáveis selecionadas.

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

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

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

Termos aninhados

É possível construir termos aninhados para o seu modelo nesse procedimento. Termos aninhados são úteis para modelar o efeito de um fator ou de uma covariável cujos valores não interagem com os níveis

de um outro fator. Por exemplo, uma cadeia de supermercados pode seguir os hábitos de consumo de seus clientes em vários locais de armazenamento. 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 interação, como termos polinomiais envolvendo a mesma covariável ou incluir vários níveis de aninhamento no termo aninhado.

Limitações. Termos aninhados têm 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. Assim, se A for um fator e X uma covariável, então especificar $A(X)$ será 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 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 convergência for obtida durante a fase de escoragem de Fisher do método híbrido antes que o número máximo de iterações de Fisher seja atingido, o algoritmo continua com o método de 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 seguintes opções 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.

Critérios de Convergência. As opções a seguir estão disponíveis

- **Convergência de Parâmetro.** Quando selecionado, o algoritmo para após uma iteração na qual a mudança absoluta ou relativa nas estimativas paramétrica é inferior ao 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 correspondentes 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 iniciar com a variável ou variáveis de divisão de arquivo na ordem especificada ao criar o Arquivo Dividido, seguido por *RowType_*, *VarName_*, *P1*, *P2*, ... conforme acima. Divisões devem ocorrer no conjunto de dados especificado na mesma ordem que no conjunto de dados original.

Nota: Os nomes de variável *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 no ranqueamento da variável, não no 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 seguintes opções 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 paramétrica; 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 paramétrica e estatística do teste e intervalos de confiança correspondentes. É possível opcionalmente exibir estimativas paramétrica exponenciado além das estimativas paramétrica bruto.
- **Matriz de covariâncias para estimativas paramétrica.** Exibe a matriz de covariâncias paramétrica estimado.
- **Matriz de correlações para estimativas paramétrica.** 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 paramétrica 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 paramétrica para cada n^{a} iteração que inicia com a 0^{a} 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 da n .
- **Teste de multiplicadores de Lagrange.** 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.

- **Entre Pares.** 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 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.

- *Distância de cook.* Uma medida do quanto os resíduos de todos os casos seriam alterados se um caso específico 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.
- *Alavancar valor.* Mede 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 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.
- **Resíduo de Pearson Padronizado.** O resíduo de Pearson multiplicado pela raiz quadrada do inverso do produto do parâmetro de escala e alavancagem l para o caso.
- **Resíduo de deviance.** A raiz quadrada da contribuição de um caso para a estatística de Deviance, com o sinal do resíduo bruto.
- **Resíduo de deviance padronizado.** O resíduo de Deviance multiplicado pela raiz quadrada do inverso do produto do parâmetro de escala e alavancagem $-l$ 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 Statistics 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_.** Utiliza valores (e rótulos de valor) *COV* (covariâncias), *CORR* (correlações), *EST* (estimativas paramétrica), *SE* (erros padrão), *SIG* (níveis de significância) e *DF* (graus de liberdade de plano 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_.** Utiliza valores *P1*, *P2*, ..., correspondentes a uma lista ordenada de todos os parâmetros de modelo estimados (exceto a escala ou parâmetros binomiais negativos), para tipos de linha *COV* ou *CORR*, com rótulos de valor correspondentes às sequências de caracteres paramétrica mostradas na tabela de Estimativas paramétrica. 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*).
- 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.

Capítulo 7. 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 ρ para elementos adjacentes, ρ^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 seguintes opções 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 seguintes opções 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 seguintes opções 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 seguintes opções 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 seguintes opções 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.

Customizada. 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 omissivo, 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 omissivo, 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 omissivo, 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 Φ^{-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.** $f(x)=[(x/(1-x))^\alpha - 1]/\alpha$, if $\alpha \neq 0$. $f(x)=\log(x)$, if $\alpha=0$. for a especificação de número necessária e deve ser um número real. Isso é apropriado apenas com a distribuição binomial.

- **Probito.** $f(x)=\Phi^{-1}(x)$, em que Φ^{-1} é a função de distribuição acumulativa normal padrão inversa. Isso é apropriado apenas com a distribuição binomial.
- **Poder.** $f(x)=x^\alpha$, if $\alpha \neq 0$. $f(x)=\log(x)$, if $\alpha=0$. for a especificação de número 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. Fatores devem ter valores válidos para um caso a ser incluído 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 construir 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.

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

Todas 3 fatores. Cria todas as interações de três fatores das variáveis selecionadas.

Todas 4 fatores. Cria todas as interações de quatro fatores das variáveis selecionadas.

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

Termos aninhados

É possível construir termos aninhados para o seu modelo nesse procedimento. Termos aninhados são úteis para modelar o efeito de um fator ou de uma covariável cujos valores não interagem com os níveis de um outro fator. Por exemplo, uma cadeia de supermercados pode seguir os hábitos de consumo de seus clientes em vários locais de armazenamento. 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. Termos aninhados têm as restrições a seguir:

- Todos os fatores dentro de uma interação devem ser exclusivos. Dessa forma, se A for um fator, então, especificar A^*A é inválido.
- Todos os fatores dentro de um efeito aninhado devem ser exclusivos. Dessa forma, se A for um fator, então, especificar $A(A)$ é 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 você conseguir presumir a passagem de dados por meio da origem, será 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 seguintes opções 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 selecionado, o algoritmo para após uma iteração na qual a mudança absoluta ou relativa nas estimativas paramétrica é inferior ao 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 correspondentes 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 iniciar com a variável ou variáveis de divisão de arquivo na ordem especificada ao criar o Arquivo Dividido, seguido por *RowType_*, *VarName_*, *P1*, *P2*, ... conforme acima. Divisões devem ocorrer no conjunto de dados especificado na mesma ordem que no conjunto de dados original.

Nota: Os nomes de variável *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 no ranqueamento da variável, não no 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 seguintes opções 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 paramétrica; 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 paramétrica e estatística do teste e intervalos de confiança correspondentes. É possível opcionalmente exibir estimativas paramétrica exponenciado além das estimativas paramétrica bruto.
- **Matriz de covariâncias para estimativas paramétrica.** Exibe a matriz de covariâncias paramétrica estimado.
- **Matriz de correlações para estimativas paramétrica.** 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 paramétrica 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 paramétrica para cada n^{a} iteração que inicia com a 0^a 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.

- **Entre Pares.** 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 Statistics 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_.** Utiliza valores (e rótulos de valor) *COV* (covariâncias), *CORR* (correlações), *EST* (estimativas paramétrica), *SE* (erros padrão), *SIG* (níveis de significância) e *DF* (graus de liberdade de plano 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_**. Utiliza valores *P1, P2, ...*, correspondentes a uma lista ordenada de todos os parâmetros de modelo estimados (exceto a escala ou parâmetros binomiais negativos), para tipos de linha *COV* ou *CORR*, com rótulos de valor correspondentes às sequências de caracteres paramétrica mostradas na tabela de Estimativas paramétrica. 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 omisso 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 omisso 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 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 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*).
- 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.

Capítulo 8. 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 usar um modelo linear generalizado misto para determinar se uma nova droga anticonvulsiva pode reduzir a taxa de crises epiléticas de um paciente. Medidas repetidas do mesmo paciente são tipicamente correlacionadas positivamente; portanto, um modelo misto 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.

Assuntos. 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 na guia Estrutura de Dados são usados para definir assuntos para a estruturas de covariâncias de resíduo e fornecem a lista dos campos possíveis para definir assuntos para estruturas de covariâncias de efeitos aleatórios sobre o 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.

Definir grupos de covariância por. Os campos categóricos especificados aqui definem conjuntos independentes paramétricas de covariância de efeitos repetidos; um para cada categoria definida pela classificação cruzada de 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. 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

Obtendo um modelo linear generalizado misto

Essa variável requer a opção de 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**.
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.

Resposta. 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 probit, 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 omissivo, 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 paramétricas de modelo é estimado para cada categoria da resposta (exceto a categoria de referência). As estimativas paramétrica 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 intercepção tradicional é substituído por um conjunto paramétricas 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 Φ^{-1} é a função de distribuição acumulativa normal padrão inversa. Isso é apropriado apenas com a distribuição binomial ou multinomial.
- **Poder.** $f(x)=x^\alpha$, if $\alpha \neq 0$. $f(x)=\log(x)$, if $\alpha=0$. for a especificação de número 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", 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 o seu modelo nesse procedimento. Termos aninhados são úteis para modelar o efeito de um fator ou de uma covariável cujos valores não interagem com os níveis de um outro fator. Por exemplo, uma cadeia de supermercados pode seguir os hábitos de consumo de seus clientes em vários locais de armazenamento. 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 os efeitos de interação, como termos polinomiais envolvendo a mesma covariável, ou incluir diversos níveis de aninhamento no termo aninhado.

Limitações. Termos aninhados têm as restrições a seguir:

- Todos os fatores dentro de uma interação devem ser exclusivos. Dessa forma, se A for um fator, então, especificar $A*A$ é inválido.
- Todos os fatores dentro de um efeito aninhado devem ser exclusivos. Dessa forma, se A for um fator, então, especificar $A(A)$ é 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 do “Bloco de Efeito Aleatório”.
2. Para editar um bloco existente, selecione o bloco que você deseja editar e clique em **Editar Bloco...** Isso abre o diálogo do “Bloco de Efeito Aleatório”.
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 arrastando para a lista de efeitos. O tipo de efeito criado depende de em que ponto de acesso você solta a seleção. Campos (flag do nominal e ordinal) categóricos são usados como fatores no modelo e campos contínuos são usados como covariáveis.

- **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 2. Descrições do botão Construtor de Efeito.

Ícone	Descrição
	Exclua termos do modelo, selecionando os termos que você deseja excluir e clicando no botão Excluir.
	Reordene os termos dentro do modelo, 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 61, clicando no botão Incluir um Termo Customizado.

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.

Definir grupos de covariância por. Os campos categóricos especificados aqui definem conjuntos independentes paramétricas 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.

Combinação de assuntos. Isso permite que você especifique assuntos de efeito aleatório a partir das combinações de pré-configuração de assuntos a partir da guia Estrutura de Dados. Por exemplo, se *Escola*, *Classe* e *Aluno* estiverem definidos como assuntos na guia Estrutura de Dados e nessa ordem, então a lista suspensa de combinação de Assuntos terá **Nenhum**, **Escola**, **Escola * Classe** e **Escola * Classe * Aluno** como opções.

Tipo de covariância de efeito aleatório. Isso especifica a estruturas de covariâncias para os resíduos. As estruturas disponíveis são:

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

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 58.

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-Estimativa. 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.
- **Grau de liberdade.** Isso especifica como os graus de liberdade são calculados para testes de significância. Escolha **Fixo para todos os testes (Método de resíduo)** se o seu tamanho da amostra for suficientemente grande ou os dados forem balanceados ou o modelo usar um tipo de covariância mais simples; por exemplo, identidade escalada ou diagonal. Este é o padrão. Escolha **Variado ao longo de testes (aproximação Satterthwaite)** se o seu tamanho da amostra for pequeno ou os dados forem desbalanceados ou o modelo usar um tipo de covariância complicado; por exemplo, não estruturado.
- **Testes de efeitos e coeficientes fixos.** Esse é o método para calcular a matriz de covariâncias de estimativas paramétrica. 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 máxima absoluta ou a mudança máxima relativa nas estimativas paramétrica for menor 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 a mudança relativa na função de log da verossimilhança for menor 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.

Este valor é utilizado 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 variáveis resposta categórica. Se a resposta for categórica, essa palavra-chave salva as probabilidades preditas das primeiras categorias n , até o valor especificado como as **Categorias máximas 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 limites superiores e inferiores do intervalo de confiança para o valor predito ou a 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 63 para obter informações adicionais.). 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 de pós-estimação do ajuste do modelo. O nome do campo padrão é *PearsonResidual*.

Confianças. Salva a confiança no valor predito para a variável de 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*.

Modelo de exportação. Isso grava o modelo em um arquivo *.zip* externo. É possível usar esse arquivo de modelo para aplicar as informações de modelo a outros arquivos de dados para propósitos de escoragem. Especifique um nome do arquivo exclusivo, válido. Se a especificação do arquivo se referir a um arquivo existente, então, o arquivo será sobrescrito.

Visualização do modelo

O procedimento cria um objeto Modelo no Visualizador. Ativando (clcando duas vezes em) esse objeto, você ganha 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 uma alternativa para o objeto Modelo, é possível gerar tabelas e gráficos dinâmicos selecionando **Tabelas e gráficos dinâmicos** no grupo de Exibições 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 no log da verossimilhança -2 (Restrito). 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 no log da verossimilhança -2 . Valores menores indicam melhores modelos. O BIC também "penaliza" modelos sobreparametrizados (por exemplo, modelos complexos com um número grande de entradas), 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. Este é o padrão.
- **Contagens de célula.** Isso exibe as contagens de célula nas células. O sombreado para o heat map ainda é baseado nas porcentagens de linhas.
- **Heat map.** Isso não exibe nenhum valor nas células, apenas o sombreado.
- **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.

Omisso. Se qualquer registro tiver valores omissos na resposta, ele será exibido em uma linha (**Omisso**) 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). Este 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.

Coefficientes 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. Este 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.

Capítulo 9. 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 seleciona variáveis *A*, *B* e *C* na lista de Fatores e, em seguida, seleciona **Interação** a partir da lista suspensa Construir Termos. O modelo resultante conterá a interação de 3 vias especificada $A*B*C$, as interações bidirecionais $A*B$, $A*C$ e $B*C$ e efeitos principais para *A*, *B* e *C*. Não especifique os parentes de ordem inferior na classe geradora.

Termos de compilação

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

Todas 3 fatores. Cria todas as interações de três fatores 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.

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.

Capítulo 10. 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 e 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.

Termos de compilação

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

Todas 3 fatores. Cria todas as interações de três fatores 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.

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 usado para obter estimativas paramétrica 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 resíduo dividido por uma estimativa do seu erro padrão. Resíduos padronizados também são conhecidos como resíduos de Pearson.
- *Resíduos ajustados.* O resíduo padronizado dividido pelo 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 com sinal de uma contribuição individual para a estatística qui-quadrado de razão de verossimilhança (G quadrado), em que o sinal é o sinal do resíduos (contagem de observados menos a contagem de esperados). 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.

Capítulo 11. 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 paramétrica. 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 e 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.

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 compilação

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

Todas 3 fatores. Cria todas as interações de três fatores 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.

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 usado para obter estimativas paramétrica 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 resíduo dividido por uma estimativa do seu erro padrão. Resíduos padronizados também são conhecidos como resíduos de Pearson.
- *Resíduos ajustados.* O resíduo padronizado dividido pelo 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 com sinal de uma contribuição individual para a estatística qui-quadrado de razão de verossimilhança (G quadrado), em que o sinal é o sinal do resíduos (contagem de observados menos a contagem de esperados). 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.

Capítulo 12. 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 Eventos de Tabelas de Mortalidade para Variáveis de Status

As ocorrências do valor ou de 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.

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.
- *Log de sobrevivência.* 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.

Capítulo 13. 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.

- *Ranqueamento de log*. Um teste para comparar a igualdade das distribuições de sobrevivência. Todos os pontos de tempo são ponderados igualmente neste teste.
- *Breslow*. Um teste para comparar a igualdade das distribuições de sobrevivência. Os momentos são ponderados pelo número de casos de risco em cada momento.
- *Tarone-Ware*. Um teste para comparar a igualdade das 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 fatores. Os testes de tendência entre pares não estão disponíveis.
- *Para cada estrato*. Executa um teste separado de igualdade de todos os níveis de fatores em cada camada. Se você não tiver uma variável de estratificação, os testes não serão executados.
- *Pairwise para cada estrato*. Compara cada par distinto de níveis de fatores 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.

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.

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 acumulativa. 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*. O 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 acumulativo. 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 acumulativa de eventos quando os casos são ordenados pelos 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.
- *Log de sobrevivência.* 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.

Capítulo 14. 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 prediz a probabilidade de que o evento de interesse tenha ocorrido em um determinado tempo t para valores determinados 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 paramétrica, 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.

- *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.

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 os seus próprios valores para o gráfico, usando o grupo de controle Mudar Valor.

É possível criar gráfico de uma linha separada para cada valor de uma covariável categórica movendo essa covariável para a caixa de texto Linhas Separadas Para. Essa opção está disponível apenas para covariáveis categóricas, que são denotadas por **(Cat)** após os seus nomes a lista de Valores de Covariável Criados em Gráfico Em.

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, $Df\beta(s)$ para a regressão e o preditor linear $X*\beta$ como novas variáveis.

- *Função de sobrevivência.* O valor da função de sobrevivência acumulativa para um determinado momento. Ele é igual à probabilidade de sobrevivência para esse período de tempo.
- *Função de sobrevivência 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 representar 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.
- *$Df\beta(s)$.* Mudança estimada em um coeficiente se um caso for removido. Uma variável é salva para cada covariável no modelo final. Os $Df\beta$ s 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.

Se você estiver executando Cox com uma covariável dependente de tempo, $Df\beta(s)$ e a variável preditora linear $X*\beta$ são as únicas variáveis que é possível salvar.

Exportar Informações de Modelo para Arquivo XML. As estimativas paramétrica 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 Statistics.
- 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.

Capítulo 15. Calculando Covariáveis Dependentes de Tempo

Há determinadas situações nas quais você gostaria de calcular um modelo de Regressão de Cox, mas a suposição de riscos proporcionais não é mantida. 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. Em tais casos, é necessário usar um modelo de Regressão de Cox estendido, que permite especificar **covariáveis dependentes de tempo**.

Para esse modelo, deve-se primeiro definir a sua covariável dependente de tempo. (Várias covariáveis dependentes de tempo podem ser especificadas usando sintaxe de comando.) 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:

- Se você deseja testar a suposição de riscos proporcionais com relação a uma determinada covariável ou estimar um modelo de regressão de Cox estendido que permita riscos não proporcionais, pode fazer isso definindo a sua variável dependente de tempo como uma função da variável de tempo $T_$ e a 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 irá indicar a você se a suposição de riscos proporcionais é razoável.
- Algumas variáveis podem ter valores diferentes em períodos de tempo diferentes, mas não são sistematicamente relacionadas com o tempo. Em tais casos, é necessário definir uma **covariável dependente de tempo segmentada**, 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 a pressão sanguínea medida uma vez por semana para as quatro semanas do estudo (identificada como $BP1$ até $BP4$), será possível definir a sua 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 em parênteses será igual a 1 para qualquer caso especificado e os restantes serão todos iguais a 0. Em outras palavras, esta função significa que se o tempo for menor que uma semana, use $BP1$; se for mais de uma semana, mas menos de duas semanas, use $BP2$, e assim por diante.

Na caixa de diálogo Calcular Covariável Dependente de Tempo, é possível usar os controles de construção de função para construir a expressão para a covariável dependente de tempo ou inseri-la diretamente na área de texto Expressão para $T_COV_$. 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. A variável resultante é chamada $T_COV_$ e deve ser incluída como uma covariável em seu modelo de Regressão de Cox.

Calculando uma Covariável Dependente de Tempo

1. Nos menus, escolha:
Analisar > Sobrevivência > Cox com Covariável Dependente do Tempo ...
2. Insira uma expressão para a covariável dependente de tempo.
3. Clique em **Modelo** para continuar com a Regressão de Cox.

Nota: Assegure-se de incluir a nova variável $T_COV_$ como uma covariável no modelo de Regressão de Cox.

Consulte o tópico Capítulo 14, “Análise de Regressão de Cox”, na página 91 para obter mais informações

Regressão de Cox com Variáveis Adicionais de Covariáveis Dependentes de Tempo

O idioma da sintaxe de comando também permite que você especifique várias covariáveis dependentes de tempo. Outras variáveis de sintaxe de comando estão disponíveis para Regressão de Cox com ou sem covariáveis dependentes de tempo.

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

Capítulo 16. 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 ... 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)
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 polinominais 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 curvilínea.

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 contêm 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 excetua as somas de linha média para 0. Os 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^{th} é codificado como 0 para todas as variáveis indicadoras, exceto o i^{th} , que é codificado como 1.

Capítulo 17. 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.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_1$	$\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_4\sigma_1\rho_1\rho_2\rho_3$)
$(\sigma_2\sigma_1\rho_1$	σ_2^2	$\sigma_3\sigma_2\rho_2$	$\sigma_4\sigma_2\rho_2\rho_3$)
$(\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_3\sigma_2\rho_2$	σ_3^2	$\sigma_4\sigma_3\rho_3$)
$(\sigma_4\sigma_1\rho_1\rho_2\rho_3$	$\sigma_4\sigma_2\rho_2\rho_3$	$\sigma_4\sigma_3\rho_3$	σ_4^2)

AR(1). Essa é a estrutura autorregressiva de primeira ordem com variâncias homogêneas. A correlação entre quaisquer dois elementos é igual a ρ para elementos adjacentes, ρ^2 para elementos que são separados por um terceiro, e assim por diante. é restringido de forma que $-1 < \rho < 1$.

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2$	$\sigma^2\rho^3$)
$(\sigma^2\rho$	σ^2	$\sigma^2\rho$	$\sigma^2\rho^2$)
$(\sigma^2\rho^2$	$\sigma^2\rho$	σ^2	$\sigma^2\rho$)
$(\sigma^2\rho^3$	$\sigma^2\rho^2$	$\sigma^2\rho$	σ^2)

AR(1): Heterogêneo. Essa é a estrutura autorregressiva de primeira ordem com variâncias heterogêneas. A correlação entre quaisquer dois elementos é igual a r para elementos adjacentes, r^2 para dois elementos separados por um terceiro, e assim por diante. é restringido para estar entre -1 e 1 .

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho^2$	$\sigma_4\sigma_1\rho^3$)
$(\sigma_2\sigma_1\rho$	σ_2^2	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho^2$)
$(\sigma_3\sigma_1\rho^2$	$\sigma_3\sigma_2\rho$	σ_3^2	$\sigma_4\sigma_3\rho$)
$(\sigma_4\sigma_1\rho^3$	$\sigma_4\sigma_2\rho^2$	$\sigma_4\sigma_3\rho$	σ_4^2)

ARMA(1,1). Essa é a estrutura média móvel autorregressiva de primeira ordem. Ela tem variâncias homogêneas. A correlação entre dois elementos é igual a ϕ para elementos adjacentes, ϕ^2 para elementos separados por um terceiro, e assim por diante. ϕ e θ são os parâmetros autorregressivos e de média móvel, respectivamente, e os seus valores são restritos para estarem entre -1 e 1 , inclusive.

$(\sigma^2$	$\sigma^2\phi\rho$	$\sigma^2\phi\rho^2$	$\sigma^2\phi\rho^3$)
$(\sigma^2\phi\rho$	σ^2	$\sigma^2\phi\rho$	$\sigma^2\phi\rho^2$)
$(\sigma^2\phi\rho^2$	$\sigma^2\phi\rho$	σ^2	$\sigma^2\phi\rho$)
$(\sigma^2\phi\rho^3$	$\sigma^2\phi\rho^2$	$\sigma^2\phi\rho$	σ^2)

Simetria Composta. Essa estrutura tem variância constante e covariância constante.

$(\sigma^2 + \sigma_1^2$	σ_1	σ_1	σ_1)
$(\sigma_1$	$\sigma^2 + \sigma_1^2$	σ_1	σ_1)
$(\sigma_1$	σ_1	$\sigma^2 + \sigma_1^2$	σ_1)
$(\sigma_1$	σ_1	σ_1	$\sigma^2 + \sigma_1^2$)

Simetria Composta: Métrica de Correlação. Essa estruturas de covariâncias tem variâncias homogêneas e correlações homogêneas entre elementos.

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho)$
$(\sigma^2\rho$	σ^2	$\sigma^2\rho$	$\sigma^2\rho)$
$(\sigma^2\rho$	$\sigma^2\rho$	σ^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.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho$	$\sigma_4\sigma_1\rho)$
$(\sigma_2\sigma_1\rho$	σ_2^2	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho)$
$(\sigma_3\sigma_1\rho$	$\sigma_3\sigma_2\rho$	σ_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	σ_2^2	0	0)
0	0	σ_3^2	0)
0	0	0	$\sigma_4^2)$

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.

$(\lambda_1^2 + d$	$\lambda_2\lambda_1$	$\lambda_3\lambda_1$	$\lambda_4\lambda_1)$
$(\lambda_2\lambda_1$	$\lambda_2^2 + d$	$\lambda_3\lambda_2$	$\lambda_4\lambda_2)$
$(\lambda_3\lambda_1$	$\lambda_3\lambda_2$	$\lambda_3^2 + d$	$\lambda_4\lambda_3)$
$(\lambda_4\lambda_1$	$\lambda_4\lambda_2$	$\lambda_4\lambda_3$	$\lambda_4^2 + d)$

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.

$(\lambda_1^2 + d_1$	$\lambda_2\lambda_1$	$\lambda_3\lambda_1$	$\lambda_4\lambda_1)$
$(\lambda_2\lambda_1$	$\lambda_2^2 + d_2$	$\lambda_3\lambda_2$	$\lambda_4\lambda_2)$
$(\lambda_3\lambda_1$	$\lambda_3\lambda_2$	$\lambda_3^2 + d_3$	$\lambda_4\lambda_3)$
$(\lambda_4\lambda_1$	$\lambda_4\lambda_2$	$\lambda_4\lambda_3$	$\lambda_4^2 + d_4)$

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.

$(\sigma_1^2$	$[\sigma_1^2 + \sigma_2^2]/2 - \lambda$	$[\sigma_1^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_1^2 + \sigma_4^2]/2 - \lambda)$
$([\sigma_1^2 + \sigma_2^2]/2 - \lambda$	σ_2^2	$[\sigma_2^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_4^2]/2 - \lambda)$
$([\sigma_1^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_3^2]/2 - \lambda$	σ_3^2	$[\sigma_3^2 + \sigma_4^2]/2 - \lambda)$
$([\sigma_1^2 + \sigma_4^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_4^2]/2 - \lambda$	$[\sigma_3^2 + \sigma_4^2]/2 - \lambda$	$\sigma_4^2)$

Identidade Escalada. Essa estrutura tem variância constante. É assumido que não há nenhuma correlação entre qualquer elemento.

$(\sigma^2$	0	0	0)
-------------	---	---	----

$$\begin{pmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

Toeplitz. Essa estruturas de covariâncias tem variâncias homogêneas e correlações heterogêneas entre elementos. A correlação entre elementos adjacentes é homogênea ao longo de pares de elementos adjacentes. A correlação entre elementos separados por um terceiro é novamente homogênea, e assim por diante.

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

Toeplitz: Heterogênea. Essa estruturas de covariâncias tem variâncias heterogêneas e correlações heterogêneas entre elementos. A correlação entre elementos adjacentes é homogênea ao longo de pares de elementos adjacentes. A correlação entre elementos separados por um terceiro é novamente homogênea, e assim por diante.

$$\begin{pmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

Não estruturada. Essa é um matriz de covariâncias completamente geral.

$$\begin{pmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

Não estruturada: Métrica de Correlação. Essa estruturas de covariâncias tem variâncias heterogêneas e correlações heterogêneas.

$$\begin{pmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

Componentes de Variância. Essa estrutura atribui uma estrutura de identidade escalada (ID) a cada um dos efeitos aleatórios especificados.

Avisos

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

LICENÇA DE COPYRIGHT:

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

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

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

Marcas comerciais

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

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

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

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

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

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

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

Índice Remissivo

A

amplitude múltipla de
 Ryan-Einot-Gabriel-Welsch
 no GLM Multivariate 7
 no GLM Repeated Measures 16
análise de covariância
 no GLM Multivariado 3
Análise de log linear de seleção do
 modelo 71
 definindo intervalos de fator 72
 modelos 72
 opções 72
 recursos adicionais do comando 73
Análise de log linear geral
 contrastes 75
 covariáveis de célula 75
 critérios 76
 distribuição de contagens de
 células 75
 especificação de modelo 76
 estruturas de célula 75
 fatores 75
 gráficos 76
 intervalos de confiança 76
 opções de exibição 76
 recursos adicionais do comando 77
 resíduos 77
 salvando valores preditos 77
 salvando variáveis 77
análise de probit
 modelos lineares generalizados
 mistos 57
análise de sobrevivência
 em Kaplan-Meier 87
 em Regressão de Cox 91
 em Tabelas de Mortalidade 83
 Regressão de Cox dependente de
 tempo 95
análise de variância
 em Componentes de Variância 22
 em modelos lineares generalizados
 mistos 57
Análise Log-Linear Logit 79
 contrastes 79
 covariáveis de célula 79
 critérios 81
 distribuição de contagens de
 células 79
 especificação de modelo 80
 estruturas de célula 79
 fatores 79
 gráficos 81
 intervalos de confiança 81
 opções de exibição 81
 resíduos 81
 salvando variáveis 81
 valores preditos 81
análise loglinear 71
 Análise de log linear geral 75
 Análise Log-Linear Logit 79

análise loglinear (*continuação*)
 em modelos lineares generalizados
 mistos 57
ANOVA
 no GLM Multivariado 3
 no GLM Repeated Measures 11
ANOVA multivariado 3

B

Bonferroni
 no GLM Multivariate 7
 no GLM Repeated Measures 16

C

C de Dunnett
 no GLM Multivariate 7
 no GLM Repeated Measures 16
casos censurados
 em Kaplan-Meier 87
 em Regressão de Cox 91
 em Tabelas de Mortalidade 83
categoria de referência
 em Equações de Estimativa
 Generalizada 49, 50
 em Modelos Lineares
 Generalizados 36
classe geradora
 em Análise Log-Linear de Seleção de
 Modelo 72
Componentes de variância 21
 model 22
 opções 22
 recursos adicionais do comando 24
 salvando resultados 24
contrastes
 em Análise Log-Linear Geral 75
 em Análise Log-Linear Logit 79
 em Regressão de Cox 92
convergência do log da verossimilhança
 em Equações de Estimativa
 Generalizada 51
 em Modelos Lineares
 Generalizados 38
 nos Modelos lineares mistos 29
Convergência hessiana
 em Equações de Estimativa
 Generalizada 51
 em Modelos Lineares
 Generalizados 38
convergência paramétrica
 em Equações de Estimativa
 Generalizada 51
 em Modelos Lineares
 Generalizados 38
 nos Modelos lineares mistos 29
covariáveis
 em Regressão de Cox 92

covariáveis de sequência de caracteres
 em Regressão de Cox 92
covariáveis dependentes de tempo
 segmentadas
 em Regressão de Cox 95

D

decomposição hierárquica 5, 14
 em Componentes de Variância 23
desvio padrão
 no GLM Multivariate 9
 no GLM Repeated Measures 18
diferença menos significativa
 no GLM Multivariate 7
 no GLM Repeated Measures 16
diferença significativa honesta de Tukey
 no GLM Multivariate 7
 no GLM Repeated Measures 16
Distância de Cook
 em Modelos Lineares
 Generalizados 42
 no GLM 8
 no GLM Repeated Measures 18
distribuição binomial
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33
distribuição binomial negativa negativa
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33
Distribuição de Poisson
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33
Distribuição de Tweedie
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33
distribuição gama
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33
Distribuição Gaussiana inversa
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33
distribuição multinomial
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

distribuição normal
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

E

efeitos aleatórios
 nos Modelos lineares mistos 28

efeitos fixos
 nos Modelos lineares mistos 27

eliminação backward
 em Análise Log-Linear de Seleção de
 Modelo 71

Equações de estimação generalizadas
 valores iniciais 52

Equações de Estimativa Generalizada 45
 categoria de referência para resposta
 binária 49
 critérios de estimação 51
 especificação de modelo 50
 estatísticas 53
 exportação do modelo 55
 médias marginais estimadas 54
 opções para fatores categóricos 50
 preditores 49
 resposta 49
 salvar variáveis no conjunto de dados
 ativo 55
 tipo de modelo 47

erro padrão
 no GLM 8
 no GLM Multivariate 9
 no GLM Repeated Measures 18

escoragem
 nos Modelos lineares mistos 29

estatística Wald
 em Análise Log-Linear Geral 75
 em Análise Log-Linear Logit 79

estatísticas descritivas
 em Equações de Estimativa
 Generalizada 53
 em Modelos Lineares
 Generalizados 39
 no GLM Multivariate 9
 no GLM Repeated Measures 18
 nos Modelos lineares mistos 29

estimação de máxima verossimilhança
 em Componentes de Variância 22

estimação de máxima verossimilhança
 restrita
 em Componentes de Variância 22

estimativas de efeito-tamanho
 no GLM Multivariate 9
 no GLM Repeated Measures 18

estimativas de potência
 no GLM Multivariate 9
 no GLM Repeated Measures 18

estimativas paramétrica
 em Análise Log-Linear de Seleção de
 Modelo 72
 em Análise Log-Linear Geral 75
 em Análise Log-Linear Logit 79
 em Equações de Estimativa
 Generalizada 53

estimativas paramétrica (*continuação*)
 em Modelos Lineares
 Generalizados 39
 no GLM Multivariate 9
 no GLM Repeated Measures 18
 nos Modelos lineares mistos 29

estruturas de covariâncias 101
 nos Modelos lineares mistos 101

eta quadrado
 no GLM Multivariate 9
 no GLM Repeated Measures 18

etapa pela metade
 em Equações de Estimativa
 Generalizada 51
 em Modelos Lineares
 Generalizados 38
 nos Modelos lineares mistos 29

F

F múltiplo de Ryan-Einot-Gabriel-Welsch
 no GLM Multivariate 7
 no GLM Repeated Measures 16

fatores
 no GLM Repeated Measures 13

frequências
 em Análise Log-Linear de Seleção de
 Modelo 72

função de ligação
 modelos lineares generalizados
 mistos 58

função de ligação binomial negativa
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação de Cauchit acumulativa
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação de complemento de log
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação de identidade
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação de log
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação de poder
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação de poder de chances
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação de probito
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação de probito acumulativa
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação log-log
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação log-log acumulativa
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação log-log negativa
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação log-log negativa
 acumulativa
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação logit
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de ligação logit acumulativa
 em equações de estimativa
 generalizada 47
 em modelos lineares
 generalizados 33

função de sobrevivência
 em Tabelas de Mortalidade 83

função estimável geral
 em Equações de Estimativa
 Generalizada 53
 em Modelos Lineares
 Generalizados 39

G

GLM
 salvando matrizes 8
 salvando variáveis 8

GLM multivariado 3

GLM Multivariado 10

GLM Multivariate 3
 covariáveis 3
 diagnóstico 9
 exibição 9
 fatores 3
 gráficos de perfil 6
 lista de variáveis dependentes, 3
 médias marginais estimadas 9
 opções 9
 testes post hoc 7

GLM Repeated Measures 11

GLM Repeated Measures (*continuação*)
 definir fatores 13
 diagnóstico 18
 exibição 18
 gráficos de perfil 16
 médias marginais estimadas 18
 model 14
 opções 18
 recursos adicionais do comando 19
 salvando variáveis 18
 testes post hoc 16

GLOR
 em Análise Log-Linear Geral 75
 gráficos
 em Análise Log-Linear Geral 76
 em Análise Log-Linear Logit 81
 gráficos de dispersão versus nível
 no GLM Multivariate 9
 no GLM Repeated Measures 18
 gráficos de perfil
 no GLM Multivariate 6
 no GLM Repeated Measures 16
 gráficos de probabilidade normal
 em Análise Log-Linear de Seleção de
 Modelo 72
 gráficos de resíduos
 no GLM Multivariate 9
 no GLM Repeated Measures 18

GT2 de Hochberg
 no GLM Multivariate 7
 no GLM Repeated Measures 16

H

histórico de iteração
 em Equações de Estimativa
 Generalizada 53
 em Modelos Lineares
 Generalizados 39
 nos Modelos lineares mistos 29

I

informações a priori de efeito aleatório
 em Componentes de Variância 22
 informações de modelo
 em Equações de Estimativa
 Generalizada 53
 em Modelos Lineares
 Generalizados 39
 informações de nível do fator
 nos Modelos lineares mistos 29
 intervalos de confiança
 em Análise Log-Linear Geral 76
 em Análise Log-Linear Logit 81
 no GLM Multivariate 9
 no GLM Repeated Measures 18
 nos Modelos lineares mistos 29
 iterações
 em Análise Log-Linear de Seleção de
 Modelo 72
 em Equações de Estimativa
 Generalizada 51
 em Modelos Lineares
 Generalizados 38

K

Kaplan-Meier 87
 comparar níveis de fator 88
 definindo eventos 88
 estatísticas 87, 89
 example 87
 gráficos 89
 quartis 89
 recursos adicionais do comando 89
 salvando novas variáveis 88
 tabelas de sobrevivência 89
 tempo de sobrevivência média e
 mediana 89
 tendência linear para níveis do
 fator 88
 variáveis de status de
 sobrevivência 88

L

LSD de Fisher
 no GLM Multivariate 7
 no GLM Repeated Measures 16

M

matriz de correlações
 em Equações de Estimativa
 Generalizada 53
 em Modelos Lineares
 Generalizados 39
 nos Modelos lineares mistos 29
 matriz de covariância paramétrica
 nos Modelos lineares mistos 29
 matriz de covariância residual
 nos Modelos lineares mistos 29
 matriz de covariâncias
 em Equações de Estimativa
 Generalizada 51, 53
 em Modelos Lineares
 Generalizados 38, 39
 no GLM 8
 nos Modelos lineares mistos 29
 matriz de covariâncias de efeito aleatório
 nos Modelos lineares mistos 29
 matriz dos coeficientes de contraste
 em Equações de Estimativa
 Generalizada 53
 em Modelos Lineares
 Generalizados 39
 matriz L.
 em Equações de Estimativa
 Generalizada 53
 em Modelos Lineares
 Generalizados 39
 médias marginais estimadas
 em Equações de Estimativa
 Generalizada 54
 em Modelos Lineares
 Generalizados 41
 no GLM Multivariate 9
 no GLM Repeated Measures 18
 nos Modelos lineares mistos 30
 médias observadas
 no GLM Multivariate 9
 no GLM Repeated Measures 18

método Newton-Raphson
 em Análise Log-Linear Geral 75
 em Análise Log-Linear Logit 79

MINQUE
 em Componentes de Variância 22
 modelo de risco proporcional
 em Regressão de Cox 91
 modelo linear generalizado
 em modelos lineares generalizados
 mistos 57
 modelo linear geral
 modelos lineares generalizados
 mistos 57
 modelos customizados
 em Análise Log-Linear de Seleção de
 Modelo 72
 em Componentes de Variância 22
 no GLM Repeated Measures 14
 modelos de vários níveis
 modelos lineares generalizados
 mistos 57
 modelos fatoriais completos
 em Componentes de Variância 22
 no GLM Repeated Measures 14
 modelos hierárquicos
 modelos lineares generalizados
 mistos 57
 Modelos lineares generalizados 33
 categoria de referência para resposta
 binária 36
 critérios de estimação 38
 distribution 33
 especificação de modelo 37
 estatísticas 39
 exportação de modelo 43
 função de ligação 33
 médias marginais estimadas 41
 opções para fatores categóricos 37
 preditores 36
 resposta 35
 salvar variáveis no conjunto de dados
 ativo 42
 tipos de modelo 33
 valores iniciais 39
 modelos lineares generalizados
 mistos 57
 bloco de efeito aleatório 62
 coeficientes fixos 68
 covariâncias de efeito aleatório 69
 distribuição de resposta 58
 efeitos aleatórios 62
 efeitos fixos 60, 68
 estrutura de dados 67
 exportação do modelo 66
 função de ligação 58
 médias estimadas 69, 70
 médias marginais estimadas 65
 offset 63
 parâmetros de covariância 69
 ponderação de análise 63
 predito por observado 67
 salvar campos 66
 sumarização do modelo 67
 tabela de classificação 67
 termos customizados 61
 visualização do modelo 66
 modelos log-lineares hierárquicos 71

modelos logit de multinômio 79
 modelos longitudinais
 modelos lineares generalizados mistos 57
 modelos mistos
 linear 25
 modelos lineares generalizados mistos 57
 Modelos mistos lineares 25, 101
 construir termos 27
 critérios de estimação 29
 efeitos aleatórios 28
 efeitos fixos 27
 estruturas de covariâncias 101
 médias marginais estimadas 30
 model 29
 recursos adicionais do comando 31
 salvar variáveis 30
 termos de interação 27
 modelos saturados
 em Análise Log-Linear de Seleção de Modelo 72

N

Newman-Keuls
 no GLM Multivariate 7
 no GLM Repeated Measures 16

P

parâmetro de escala
 em Equações de Estimativa Generalizada 51
 em Modelos Lineares Generalizados 38
 Pontuação de Fisher
 nos Modelos lineares mistos 29
 produtos cruzados
 matrizes de hipótese e erros 9
 proporção fracionária
 em Análise Log-Linear Geral 75

Q

qualidade do ajuste
 em Equações de Estimativa Generalizada 53
 em Modelos Lineares Generalizados 39

R

R-E-G-W F
 no GLM Multivariate 7
 no GLM Repeated Measures 16
 R-E-G-W Q
 no GLM Multivariate 7
 no GLM Repeated Measures 16
 razão de possibilidades de log generalizada
 em Análise Log-Linear Geral 75
 Regressão de Cox 91
 contrastes 92
 covariáveis 91

Regressão de Cox (*continuação*)
 covariáveis categóricas 92
 covariáveis de sequência de caracteres 92
 covariáveis dependentes de tempo 95, 96
 definir evento 94
 DfBeta(s) 93
 entrada e remoção stepwise 93
 estatísticas 91, 93
 example 91
 função de risco 93
 função de sobrevivência 93
 funções da linha de base 93
 gráficos 92
 iterações 93
 recursos adicionais do comando 94
 resíduo parcial 93
 salvando novas variáveis 93
 variável de status de sobrevivência 94
 regressão de Poisson
 em Análise Log-Linear Geral 75
 modelos lineares generalizados mistos 57
 regressão logística multinomial
 modelos lineares generalizados mistos 57
 regressão multivariada 3
 regressão regressão logística
 modelos lineares generalizados mistos 57
 resíduos
 em Análise Log-Linear de Seleção de Modelo 72
 em Análise Log-Linear Geral 77
 em Análise Log-Linear Logit 81
 em Equações de Estimativa Generalizada 55
 em Modelos Lineares Generalizados 42
 nos Modelos lineares mistos 30
 resíduos de deviance
 em Modelos Lineares Generalizados 42
 resíduos de Pearson
 em Equações de Estimativa Generalizada 55
 em Modelos Lineares Generalizados 42
 resíduos de probabilidade
 em Modelos Lineares Generalizados 42
 resíduos excluídos
 no GLM 8
 no GLM Repeated Measures 18
 resíduos não padronizados
 no GLM 8
 no GLM Repeated Measures 18
 resíduos padronizados
 no GLM 8
 no GLM Repeated Measures 18

S

separação
 em Equações de Estimativa Generalizada 51
 em Modelos Lineares Generalizados 38
 soma dos quadrados 5, 14
 somas de quadrados
 em Componentes de Variância 23
 matrizes de hipótese e erros 9
 nos Modelos lineares mistos 27
 SSCP
 no GLM Multivariate 9
 no GLM Repeated Measures 18
 SSCP de resíduo
 no GLM Multivariate 9
 no GLM Repeated Measures 18
 Student-Newman-Keuls
 no GLM Multivariate 7
 no GLM Repeated Measures 16
 sumarização de processamento de caso
 em Equações de Estimativa Generalizada 53
 em Modelos Lineares Generalizados 39

T

T2 de Tamhane
 no GLM Multivariate 7
 no GLM Repeated Measures 16
 T3 de Dunnett
 no GLM Multivariate 7
 no GLM Repeated Measures 16
 tabelas de contingência
 em Análise Log-Linear Geral 75
 Tabelas de mortalidade 83
 comparar níveis de fator 84
 estatísticas 83
 example 83
 função de sobrevivência 83
 gráficos 84
 recursos adicionais do comando 85
 suprimindo exibição da tabela 84
 taxa de risco 83
 teste de Wilcoxon (Gehan) 84
 variáveis de fator 84
 variáveis de status de sobrevivência 84
 tabulação cruzada
 em Análise Log-Linear de Seleção de Modelo 71
 taxa de risco
 em Tabelas de Mortalidade 83
 termos aninhados
 em Equações de Estimativa Generalizada 50
 em Modelos Lineares Generalizados 37
 nos Modelos lineares mistos 27
 termos de compilação 4, 14, 22, 72, 76, 80
 termos de interação 4, 14, 22, 72, 76, 80
 nos Modelos lineares mistos 27
 teste b de Tukey
 no GLM Multivariate 7

teste b de Tukey (<i>continuação</i>)		valores preditos (<i>continuação</i>)	
no GLM Repeated Measures	16	em Análise Log-Linear Logit	81
teste de amplitude múltipla de Duncan		nos Modelos lineares mistos	30
no GLM Multivariate	7	valores preditos fixos	
no GLM Repeated Measures	16	nos Modelos lineares mistos	30
teste de Breslow		valores preditos ponderados	
em Kaplan-Meier	88	no GLM	8
teste de comparação entre pares de Gabriel		no GLM Repeated Measures	18
no GLM Multivariate	7	variáveis de assunto	
no GLM Repeated Measures	16	nos Modelos lineares mistos	26
teste de comparações pairwise de Games-Howell		variáveis de medidas repetidas	
no GLM Multivariate	7	em Modelos Lineares Mistos	26
no GLM Repeated Measures	16	visualização do modelo	
teste de esfericidade de Bartlett		em modelos lineares generalizados	
no GLM Multivariate	9	mistos	66
Teste de esfericidade de Mauchly			
no GLM Repeated Measures	18		
teste de Gehan			
em Tabelas de Mortalidade	84		
teste de Levene			
no GLM Multivariate	9		
no GLM Repeated Measures	18		
teste de ranqueamento de log			
em Kaplan-Meier	88		
teste de Scheffé			
no GLM Multivariate	7		
no GLM Repeated Measures	16		
teste de Tarone-Ware			
em Kaplan-Meier	88		
teste de Wilcoxon			
em Tabelas de Mortalidade	84		
teste dos multiplicadores de Lagrange			
em Modelos Lineares			
Generalizados	39		
teste M de Box			
no GLM Multivariate	9		
teste paramétrica de covariância			
nos Modelos lineares mistos	29		
teste t			
no GLM Multivariate	9		
no GLM Repeated Measures	18		
teste t de Dunnett			
no GLM Multivariate	7		
no GLM Repeated Measures	16		
teste t de Sidak			
no GLM Multivariate	7		
no GLM Repeated Measures	16		
teste t de Waller-Duncan			
no GLM Multivariate	7		
no GLM Repeated Measures	16		
testes de homogeneidade de variância			
no GLM Multivariate	9		
no GLM Repeated Measures	18		
tolerância a singularidade			
nos Modelos lineares mistos	29		

V

valores de ponto de alavanca	
em Modelos Lineares	
Generalizados	42
no GLM	8
no GLM Repeated Measures	18
valores preditos	
em Análise Log-Linear Geral	77



Impresso no Brasil