

*IBM SPSS Regression 31*



**Observação**

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

**Informações sobre o produto**

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

© Copyright International Business Machines Corporation .

# Índice

<b>Capítulo 1. Regressão.....</b>	<b>1</b>
Escolhendo um procedimento para Regressão logística binária.....	1
Regressão logística.....	2
Regra do conjunto de regressão logística.....	3
Métodos de seleção de variáveis de regressão logística.....	3
Definir variáveis categóricas de regressão logística.....	3
Regressão logística - Salvar novas variáveis.....	4
Opções de regressão logística.....	5
Recursos adicionais do comando LOGISTIC REGRESSION.....	5
Regressão logística multinomial.....	6
Regressão logística multinomial.....	6
Categoria de referência de regressão logística multinomial.....	7
Estatísticas de regressão logística multinomial.....	7
Critérios de Regressão logística multinomial.....	8
Opções de regressão logística multinomial.....	8
Regressão logística multinomial - Salvar.....	9
Recursos adicionais do comando NOMREG.....	10
Regressão probito.....	10
Definir intervalo de Análise de probito.....	11
Opções de análise de probit.....	11
Recursos adicionais do comando PROBIT.....	11
Regressão de quantil.....	12
Regressão de quantil: Critérios.....	13
Regressão de quantil: Modelo.....	14
Regressão de quantil: Exibição.....	15
Regressão de quantil: Salvar.....	16
Regressão de quantil: Exportação.....	17
Regressão não linear.....	18
Lógica condicional (Regressão não linear).....	18
Parâmetros de regressão não linear.....	19
Modelos comuns de regressão não linear.....	19
Função de perda de regressão não linear.....	20
Restrições paramétrica de regressão não linear.....	20
Regressão não linear - Salvar novas variáveis.....	21
Opções de regressão não linear.....	21
Interpretando resultados de regressão não linear.....	21
Recursos adicionais do comando NLR.....	22
Estimação Ponderada.....	22
Opções de estimação ponderada.....	23
Recursos adicionais do comando WLS.....	23
Regressão por quadrados mínimos de dois estágios.....	24
Opções de regressão de quadrados mínimos de dois estágios.....	25
Recursos adicionais do comando 2SLS.....	25
Esquemas de codificação de variável categórica.....	25
Desvio.....	25
Simples.....	26
Helmert.....	26
Diferença.....	26
Polinomial.....	27
Repetido.....	27
Especial.....	27

Indicador.....	28
Regressão de Kernel Ridge.....	28
Parâmetros de kernel.....	30
Regressão de Kernel Ridge: opções.....	30
Modelos De Tempo De Falha Acelerada Paramétrica.....	31
Modelos De Tempo De Falha Acelerada Paramétrica: Critérios.....	32
Modelos De Tempo De Falha Acelerada Paramétrica: Modelo.....	32
Modelos De Tempo De Falha Acelerada Paramétrica: Estimativa.....	33
Modelos De Tempo De Falha Acelerada Paramétrica: Imprimir.....	34
Modelos De Tempo De Falha Acelerada Paramétrica: Prever.....	35
Modelos De Tempo De Falha Acelerada Paramétrica: Plot.....	36
Modelos De Tempo De Falha Acelerada Paramétrica: Exportação.....	36
Sobrevivência AFT Definir Eventos para Variáveis de Status.....	37
Modelos de Tempo de Falha Acelerada Paramétrica: Selecionar Categoria.....	37
Modelos de fragilidade compartilhados paramétricos.....	38
Modelos de Fragilidade Compartilhados Paramétricos: Critérios.....	39
Modelos de fragilidade compartilhados paramétricos: Modelo.....	39
Modelos Paramétricos de Fragilidade Compartilhados: Estimativa.....	40
Modelos de fragilidade compartilhados paramétricos: Impressão.....	41
Modelos de fragilidade compartilhados paramétricos: prever.....	42
Modelos de Fragilidade Compartilhados Paramétricos: Plot.....	43
Modelos paramétricos compartilhados de fragilidade: Exportação.....	43
Modelos de fragilidade compartilhados paramétricos: definir eventos.....	44
Modelos de fragilidade compartilhados paramétricos-Exemplos.....	44
Modelos paramétricos compartilhados de fragilidade-um estudo de caso para dados recuperados.....	46
<b>Avisos.....</b>	<b>49</b>
Marcas comerciais.....	50
<b>Índice remissivo.....</b>	<b>53</b>

---

# Capítulo 1. Regressão

Os seguintes recursos de regressão estão incluídos em SPSS Statistics Edição Padrão ou a opção Regressão.

## Escolhendo um procedimento para Regressão logística binária

---

Os modelos de regressão logística binários podem ser ajustados usando o procedimento Regressão logística e o procedimento Regressão logística multinomial. Cada procedimento possui opções não disponíveis nos outros. Uma distinção teórica importante é que o procedimento Regressão logística produz todas as previsões, resíduos, estatísticas de influência e testes de qualidade do ajuste usando dados no nível de caso individual, independentemente de como os dados são inseridos e se o número de padrões de covariável é ou não menor que o número total de casos, enquanto o procedimento Regressão logística multinomial agrega casos internamente para formar subpopulações com padrões de covariável idênticos para os preditores, produzindo previsões, resíduos e testes de qualidade do ajuste com base nessas subpopulações. Se todos os preditores forem categóricos ou os preditores contínuos usarem somente um número limitado de valores — para que haja vários casos em cada padrão de covariável distinto — a abordagem da subpopulação pode produzir testes de qualidade do ajuste válidos e resíduos informativos, enquanto a abordagem de nível de caso individual não pode.

### Regressão logística

Fornece os seguintes recursos exclusivos:

- Teste de qualidade do ajuste de Hosmer-Lemeshow para o modelo
- Análises stepwise
- Contrastes para definir a parametrização de modelo
- Pontos de corte alternativos para classificação
- Gráficos de classificação
- Modelo ajustado em um conjunto de casos para um conjunto de casos eficaz
- Salva previsões, resíduos e estatísticas de influência

### Regressão logística multinomial

Fornece os seguintes recursos exclusivos:

- Testes qui-quadrado de Pearson e deviance para qualidade do ajuste do modelo
- Especificação de subpopulações para agrupamento de dados para testes de qualidade do ajuste
- Listagem de contagens, contagens previstas e resíduos por subpopulações
- Correção de estimativas de variância para sobredispersão
- Matriz de covariâncias das estimativas de parâmetros
- Testes de combinações lineares paramétricas
- Especificação explícita de modelos aninhados
- Ajustar modelos de regressão logística condicionais correspondentes a 1-1 usando variáveis diferenciadas

### Notas:

- Dois desses procedimentos ajustam um modelo para dados binários que é um modelo linear generalizado com uma distribuição binomial e função de ligação logit. Se uma função de ligação diferente for mais apropriada para seus dados, você deverá usar o procedimento Modelos lineares generalizados.
- Se você tiver medidas repetidas de dados binários ou registros que, de outra forma, são correlacionados, deverá considerar os procedimentos de Modelos Lineares Generalizados Mistos ou de Equações de Estimativa Generalizada.

## Regressão logística

---

A regressão logística é útil para situações nas quais você deseja poder prever a presença ou ausência de uma característica ou resultado com base em valores de um conjunto de variáveis preditoras. É semelhante a um modelo de regressão linear, mas é adequado para modelos em que a variável dependente é dicotômica. Os coeficientes de regressão logística podem ser usados para estimar razões de chances para cada uma das variáveis independentes no modelo. A regressão logística é aplicável a um intervalo mais amplo de situações de pesquisa do que de análise discriminante.

**Exemplo.** Quais características de estilo de vida são fatores de risco para doença coronariana (CHD)? Em uma amostra de pacientes em que foram registrados o status de fumante, dieta, exercícios, uso de bebidas alcoólicas e status de CHD, foi possível construir um modelo usando as quatro variáveis de estilo de vida para prever a presença ou ausência de CHD em uma amostra de pacientes. O modelo pode então ser usado para derivar estimativas das razões de chances para cada fator para informar, por exemplo, quanto mais probabilidades têm os fumantes de desenvolver CHD do que os não fumantes.

**Estatísticas.** Para cada análise: total de casos, casos selecionados, casos válidos. Para cada variável categórica: codificação paramétrica. Para cada passo: variáveis inseridas ou removidas, histórico de iteração,  $-2 \log$  da verossimilhança, qualidade do ajuste, estatísticas de qualidade de ajuste de Hosmer-Lemeshow, qui-quadrado de modelo, qui-quadrado de melhoria, tabela de classificação, correlações entre variáveis, gráfico de grupos observados e de probabilidades preditas, qui-quadrado de resíduo. Para cada variável na equação: coeficiente ( $B$ ), erro padrão de  $B$ , estatística de Wald, razão de chances estimada ( $\exp(B)$ ), intervalo de confiança para  $\exp(B)$ , log da verossimilhança se o termo tiver sido removido do modelo. Para cada variável não na equação: estatística de escore. Para cada caso: grupo observado, probabilidade predita, grupo predito, resíduo, resíduo padronizado.

**Métodos.** É possível estimar modelos usando entrada de bloco de variáveis ou qualquer um dos seguintes métodos stepwise: forward condicional, forward LR, forward Wald, backward condicional, backward LR ou backward Wald.

### Considerações de dados de Regressão logística

**Dados.** A variável dependente deve ser dicotômica. As variáveis independentes podem ser de nível de intervalo ou categóricas; se categóricas, elas devem ser simulado ou codificadas por indicador (há uma opção no procedimento para recodificar variáveis categóricas automaticamente).

**Suposições.** A regressão logística não depende de suposições distributivas da mesma forma que a análise discriminante depende. No entanto, sua solução pode ser mais estável se seus preditores tiverem uma distribuição normal multivariada. Além disso, assim como com outras formas de regressão, a multicolinearidade entre os preditores pode conduzir a estimativas parciais e a erros padrão desserializados. O procedimento é mais eficiente quando a associação ao grupo é de fato uma variável categórica; se a associação ao grupo for baseada em valores de uma variável contínua (por exemplo, "QI alto" versus "QI baixo"), deve-se considerar o uso de regressão linear para tirar vantagem das informações mais importantes oferecidas pela própria variável contínua.

**Procedimentos relacionados.** Use o procedimento Gráfico de dispersão para triar seus dados para multicolinearidade. Se as suposições de normalidade multivariada e de matrizes de variância-covariância iguais forem atendidas, será possível obter uma solução mais rápida usando o procedimento Análise discriminante. Se todas as suas variáveis preditoras forem categóricas, também será possível usar o procedimento Log linear. Se sua variável dependente for contínua, use o procedimento Regressão linear. É possível usar o procedimento Curva ROC para probabilidades de gráfico salvas com o procedimento Regressão logística.

### Obtendo uma análise de regressão logística

1. Nos menus, escolha:

**Analisar > Regressão > Logística Binária...**

2. Selecione uma variável dependente dicotômica. Essa variável pode ser numérica ou de sequência de caracteres.

3. Selecione uma ou mais covariáveis. Para incluir termos de interação, selecione todas as variáveis envolvidas na interação e, em seguida, selecione **>a\*b>**.

Para inserir variáveis em grupos (**blocos**), selecione as covariáveis para um bloco e clique em **Avançar** para especificar um novo bloco. Repita até que todos os blocos tenham sido especificados.

Opcionalmente, é possível selecionar casos para análise. Escolha uma variável de seleção e insira os critérios de regra.

## Regra do conjunto de regressão logística

Os casos definidos pela regra de seleção são incluídos na estimação do modelo. Por exemplo, se você selecionou uma variável e **igual** e especificou um valor 5, somente os casos para os quais a variável selecionada tem um valor igual a 5 serão incluídos na estimativa do modelo.

Estatísticas e resultados de classificação são gerados para casos selecionados e não selecionados. Isso fornece um mecanismo para classificar novos casos com base nos dados existentes anteriormente, ou para particionar seus dados em subconjuntos de treinamento e teste, para executar a validação no modelo gerado.

## Métodos de seleção de variáveis de regressão logística

A seleção de método permite especificar como variáveis independentes são inseridas na análise. Usando métodos diferentes, é possível construir uma variedade de modelos de regressão a partir do mesmo conjunto de variáveis.

- *Inserir*. Um procedimento para seleção de variáveis em que todas as variáveis em um bloco são inseridas em um único passo.
- *Seleção Forward (Condicional)*. Método de seleção stepwise com o teste de entrada com base na significância da estatística de escore e com o teste de remoção com base na probabilidade de uma estatística de razão de verossimilhança com base nas estimativas paramétrica condicional.
- *Seleção Forward (Razão de Verossimilhança)*. Método de seleção stepwise com o teste de entrada com base na significância da estatística de escore e com o teste de remoção com base na probabilidade de uma estatística de razão de verossimilhança baseado nas estimativas de probabilidade parcial máximas.
- *Seleção Forward (Wald)*. Método de seleção stepwise com o teste de entrada com base na significância da estatística de escore e com o teste de remoção com base na probabilidade da estatística Wald.
- *Eliminação Backward (Condicional)*. Seleção stepwise backward. O teste de remoção é baseado na probabilidade da estatística de razão de verossimilhança com base nas estimativas paramétrica condicionais.
- *Eliminação Backward (Razão de Verossimilhança)*. Seleção stepwise backward. O teste de remoção é baseado na probabilidade da estatística de razão de verossimilhança com base nas estimativas de probabilidade parcial máxima.
- *Eliminação Backward (Wald)*. Seleção stepwise backward. O teste de remoção é baseado na probabilidade da estatística Wald.

Os valores de significância em sua saída baseiam-se no ajuste de um modelo único. Portanto, os valores de significância geralmente são inválidos quando é usado um método stepwise.

Todas as variáveis independentes selecionadas são incluídas em um modelo de regressão único. Entretanto, é possível especificar métodos de entrada diferentes para subconjuntos diferentes de variáveis. Por exemplo, é possível inserir um bloco de variáveis no modelo de regressão utilizando a seleção stepwise e um segundo bloco utilizando a seleção forward. Para incluir um segundo bloco de variáveis no modelo de regressão, clique em **Avançar**.

## Definir variáveis categóricas de regressão logística

É possível especificar detalhes de como o procedimento Regressão logística tratará variáveis categóricas:

**Covariáveis.** Contém uma lista de todas as covariáveis especificadas na caixa de diálogo principal, por si mesmas ou como parte de uma interação, em qualquer estrato. 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 última categoria) é comparada com a próxima categoria.
- **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.

## Regressão logística - Salvar novas variáveis

É possível salvar resultados da regressão logística como novas variáveis no conjunto de dados ativo:

**Valores Preditos.** Salva valores preditos pelo modelo. As opções disponíveis são Probabilidades e Associação ao grupo.

- *Probabilidades.* Para cada caso, salva a probabilidade prevista de ocorrência do evento. Uma tabela na saída exibe o nome e o conteúdo de quaisquer novas variáveis. O "evento" é a categoria da variável dependente com o valor superior; por exemplo, se a variável dependente assumir os valores 0 e 1, a probabilidade predita da categoria 1 será salva.
- *Associação ao Grupo Predito.* O grupo com maior probabilidade posterior, com base em escores discriminantes. O grupo ao qual o modelo que prediz o caso pertence.

**Influência.** Salva valores de estatísticas que medem a influência de casos em valores preditos. As opções disponíveis são valores de Cook, de ponto de alavanca e DfBeta(s).

- *de Cook.* O análogo de regressão logística da estatística de influência 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.
- *Valor de Ponto de Alavanca.* A influência relativa de cada observação no ajuste do modelo.
- *DfBetas.* A diferença no valor beta é a mudança no coeficiente de regressão resultante da exclusão de um determinado caso. Um valor é calculado para cada termo no modelo, incluindo a constante.



**Residuais.** Salva resíduos. As opções disponíveis são Não padronizado, Logit, Estudentizado, Padronizado e Deviance.

- *Resíduos Não Padronizados.* A diferença entre um valor observado e o valor predito pelo modelo
- *Logit Residual.* O resíduo para o caso se for previsto na escala logit. O logit de resíduo é o resíduo dividido pela probabilidade predita vezes 1 menos a probabilidade predita.
- *Resíduo Estudentizado.* A mudança no desvio do modelo, se um caso for excluído.
- *Resíduos Padronizados.* O resíduo dividido por uma estimativa de seu desvio padrão Resíduos padronizados, também conhecidos como resíduos de Pearson, possuem uma média de 0 e um desvio padrão de 1.
- *Deviance.* Resíduos com base no deviance do modelo.

**Exportar informações de modelo em arquivo XML.** As estimativas paramétrica e (opcionalmente) suas covariâncias são exportadas no arquivo especificado em formato XML (PMML). É possível usar esse arquivo de modelo para aplicar as informações de modelo a outros arquivos de dados para propósitos de escoragem.

## Opções de regressão logística

É possível especificar opções para sua análise de regressão logística:

**Estatísticas e gráficos.** Permite solicitar estatísticas e gráficos. As opções disponíveis são gráficos de Classificação, qualidade do ajuste de Hosmer-Lemeshow, listagem de resíduos Casewise, Correlações de estimativas, Histórico de iteração e IC para  $\exp(B)$ . Selecione uma das alternativas no grupo Exibição para exibir estatísticas e gráficos Em cada passo ou somente para o modelo final, No último passo.

- *Estatísticas de qualidade de ajuste de Hosmer-Lemeshow.* Essa estatística de qualidade de ajuste é mais robusta do que a estatística de qualidade de ajuste tradicional usada na regressão logística, particularmente para modelos com covariáveis contínuas e estudos com tamanhos de amostra pequenos. Baseia-se casos de agrupamento em decis de risco e compara a probabilidade observada com a probabilidade esperada em cada decil.

**Probabilidade para stepwise.** Permite controlar os critérios pelos quais as variáveis são inseridas e removidas da equação. É possível especificar critérios para entrada ou remoção de variáveis.

- *Probabilidade para Stepwise.* Uma variável será inserida no modelo se a probabilidade de sua estatística de pontuação for menor que o valor de Entrada e será removida se a probabilidade for maior que o valor de Remoção Para substituir as configurações padrão, insira valores positivos para Entrada e Remoção. Entrada deve ser menor que Remoção.

**Corte de Classificação.** Permite determinar o ponto de corte para classificação de casos. Os casos com valores preditos que excederem o corte de classificação são classificados como positivos, ao passo que aqueles com valores preditos menores que o corte são classificados como negativos. Para alterar o padrão, insira um valor entre 0,01 e 0,99.

**Máximo de Iterações.** Permite mudar p número máximo de vezes que o modelo itera antes de finalizar.

**Incluir constante no modelo.** Permite indica se o modelo deve incluir um termo constante. Se desativado, o termo constante será igual a 0.

## Recursos adicionais do comando LOGISTIC REGRESSION

O idioma da sintaxe de comando também permite:

- Identificar saída casewise pelos valores ou rótulos de variáveis de uma variável.
- Controlar o espaçamento de relatórios de iteração. Em vez de imprimir estimativas paramétrica após cada iteração, é possível solicitar estimativas paramétrica após cada *enésima* iteração.
- Mudar os critérios para finalizar a iteração e verificar a redundância.
- Especificar uma lista de variáveis para listagens casewise.

- Conservar a memória mantendo os dados para cada grupo de arquivos divididos em um arquivo temporário externo durante o processamento.

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

## Regressão logística multinomial

A regressão logística multinomial é útil para situações em que você queira ser capaz de classificar assuntos com base nos valores de um conjunto de variáveis preditoras. Esse tipo de regressão é semelhante à regressão logística, mas é mais geral porque a variável dependente não está restrita a duas categorias.

**Exemplo.** Para comercializar filmes de forma mais eficiente, os estúdios de cinema querem prever o tipo de filme que um frequentador de cinema tem probabilidade de assistir. Ao executar uma Regressão logística multinomial, o estúdio pode determinar a força de influência que a idade, o gênero e o estado civil de uma pessoa têm sobre seu tipo de filme preferido. O estúdio então se concentra na campanha de publicidade de um determinado filme voltado para um grupo de pessoas que têm probabilidade de assisti-lo.

**Estatísticas.** Histórico de iteração, coeficientes paramétrica, covariância assintótica e matrizes de correlações, testes de razão de verossimilhança para modelo e efeitos parciais,  $-2 \log$  da verossimilhança. Qualidade do ajuste de qui-quadrado de Pearson e de deviance. Cox e Snell, Nagelkerke e McFadden  $R^2$ . Classificação: frequências observadas versus preditas por categoria de resposta. Tabulação cruzada: frequências observadas e preditas (com resíduos) e proporções por padrão de covariável e categoria de resposta.

**Métodos.** Um modelo logit multinomial é ajustado para o modelo fatorial completo ou um modelo especificado pelo usuário. A estimação paramétrica é executada por meio de um algoritmo de máxima verossimilhança iterativo.

### Considerações de dados de Regressão logística multinomial

**Dados.** A variável dependente deve ser categórica. As variáveis independentes podem ser fatores ou covariáveis. Em geral, fatores devem ser variáveis categóricas e covariáveis devem ser variáveis contínuas.

**Suposições.** É considerado que a razão de chances de quaisquer duas categorias é independente de todas as outras categorias de resposta. Por exemplo, se um novo produto é introduzido em um mercado, essa suposição indica que as cotas de mercado de todos os outros produtos são afetadas de forma proporcionalmente igual. Além disso, dado um padrão de covariável, as respostas são consideradas variáveis multinomiais independentes.

### Obtendo um Regressão logística multinomial

1. Nos menus, escolha:

**Analisar > Regressão > Regressão logística multinomial...**

2. Selecione uma variável dependente.
3. Os fatores são opcionais e podem ser numéricos ou categóricos.
4. As covariáveis são opcionais mas devem ser numéricas se especificadas.

## Regressão logística multinomial

Por padrão, o procedimento Regressão logística multinomial produz um modelo com os efeitos principais de fator e covariável, mas é possível especificar um modelo customizado ou solicitar a seleção de modelo stepwise com essa caixa de diálogo.

**Especificar modelo.** Um modelo dos principais efeitos contém os efeitos principais de covariável e fator, mas nenhum efeito de interação. Um modelo fatorial completo contém todos os efeitos principais e todas as interações de fator por fator. Ele não contém interações covariáveis. É possível criar um modelo

customizado para especificar subconjuntos de interações de fatores ou interações de covariáveis, ou solicitar a seleção stepwise de termos do modelo.

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

**Termos de entrada forçada.** Os termos incluídos na lista de entrada forçada são sempre incluídos no modelo.

**Termos stepwise.** Os termos incluídos na lista stepwise são incluídos no modelo, de acordo com um dos seguintes métodos stepwise selecionados pelo usuário:

- **Entrada forward.** Esse método começa sem termos stepwise no modelo. Em cada passo, o termo mais significativo é incluído no modelo até que nenhum dos termos stepwise restantes do modelo tenha uma contribuição estatisticamente significativa, se incluído no modelo.
- **Eliminação backward.** Este método começa inserindo todos os termos especificados na lista stepwise no modelo. Em cada passo, o termo stepwise menos significativo é removido do modelo até que todos os termos stepwise restantes tenham uma contribuição estatisticamente significativa para o modelo.
- **Forward stepwise.** Esse método começa com o modelo que seria selecionado pelo método de entrada forward. De lá, o algoritmo alterna entre eliminação backward nos termos stepwise no modelo e entrada forward nos termos deixados de fora do modelo. Isso continuará até que nenhum termo atenda aos critérios de entrada ou remoção.
- **Stepwise backward.** Esse método começa com o modelo que seria selecionado pelo método de eliminação backward. De lá, o algoritmo alterna entre a entrada forward nos termos que ficaram de fora do modelo e a eliminação backward nos termos stepwise no modelo. Isso continuará até que nenhum termo atenda aos critérios de entrada ou remoção.

**Incluir intercepto no modelo.** Permite incluir ou excluir um termo de intercepto para o modelo.

## Construir termos

Para os fatores e covariáveis selecionados:

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

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

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

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

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

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

## Categoria de referência de regressão logística multinomial

Por padrão, o procedimento Regressão logística multinomial torna a última categoria a categoria de referência. Esta caixa de diálogo fornece o controle da categoria de referência e a forma em que as categorias são ordenadas.

**Categoria de Referência.** Especifique a primeiro, a última ou uma categoria customizada.

**Ordem de categoria.** Em ordem crescente, o valor mais baixo define a primeira categoria e o valor mais alto define a última. Em ordem decrescente, o valor mais alto define a primeira categoria e o valor mais baixo define a última.

## Estatísticas de regressão logística multinomial

É possível especificar as seguintes estatísticas para sua Regressão logística multinomial:

**Sumarização do processamento de caso.** Esta tabela contém informações sobre as variáveis categóricas especificadas.

**Modelo.** Estatísticas para o modelo global.

- **Pseudo R-quadrado.** Imprime as estatísticas  $R^2$  de Cox e Snell, Nagelkerke e McFadden.
- **Sumarização de passo.** Essa tabela sumariza os efeitos inseridos ou removidos em cada passo em um método stepwise. Ele não será produzido, a menos que um modelo stepwise seja especificado na caixa de diálogo [Modelo](#).
- **Informações de ajuste do modelo.** Essa tabela compara os modelos ajustados e somente de interceptos ou nulos.
- **Crítérios de informações.** Essa tabela imprime o critério de informações de Akaike (AIC) e o critério de informações bayesiano (BIC) de Schwarz.
- **Probabilidades de célula.** Imprime uma tabela das frequências observadas e esperadas (com resíduos) e proporções por padrão de covariável e categoria de resposta.
- **Tabela de classificação.** Imprime uma tabela das respostas observadas versus preditas.
- **Estatísticas qui-quadrado de qualidade do ajuste.** Imprime estatísticas qui-quadrado de Pearson e de razão de verossimilhança. As estatísticas são calculadas para os padrões de covariáveis determinados por todos os fatores e covariáveis ou por um subconjunto definido pelo usuário dos fatores e covariáveis.
- **Medidas de monotonicidade.** Exibe uma tabela com informações sobre o número de pares concordantes, pares discordantes e pares empatados. O Somers' D, Gama de Goodman e Kruskal, tau-a de Kendall e o Índice C de Concordância também são exibidos nessa tabela.

**Parâmetros.** Estatísticas relacionadas aos parâmetros do modelo.

- **Estimativas.** Imprime estimativas dos parâmetros de modelo, com um nível de confiança especificado pelo usuário.
- **Teste de razão de verossimilhança.** Imprime testes de razão de verossimilhança para os efeitos parciais do modelo. O teste para o modelo global é impresso automaticamente.
- **Correlações assintóticas.** Imprime a matriz de correlações de estimativas paramétrica.
- **Covariâncias assintóticas.** Imprime a matriz de covariâncias de estimativas paramétrica.

**Definir subpopulações.** Permite selecionar um subconjunto dos fatores e covariáveis para definir os padrões de covariável usados por probabilidades de célula e os testes de qualidade do ajuste.

## Crítérios de Regressão logística multinomial

É possível especificar os seguintes critérios para sua Regressão logística multinomial:

**Iterações.** Permite especificar o número máximo de vezes que você deseja percorrer o algoritmo, o número máximo de passos na divisão do passo pela metade, as tolerâncias de convergência para mudanças no log da verossimilhança e em parâmetros, com que frequência o progresso do algoritmo iterativo é impresso e em qual iteração o processo deve começar a verificar a separação completa ou quase completa dos dados.

- **Convergência de log da verossimilhança.** A convergência será considerada se a mudança absoluta na função de log da verossimilhança for menor que o valor especificado. O critério não será utilizado se o valor for 0. Especifique um valor não negativo.
- **Convergência de Parâmetro.** A convergência será considerada se a mudança absoluta nas estimativas paramétrica for menor que esse valor. O critério não será utilizado se o valor for 0.

**Delta.** Permite que você especifique um valor não negativo menor que 1. Este valor é incluído em cada célula vazia da tabulação cruzada de categoria de resposta por padrão de covariável. Isso ajuda a estabilizar o algoritmo e evitar viés nas estimativas.

**Tolerância à singularidade.** Permite especificar a tolerância usada na verificação de singularidades.

## Opções de regressão logística multinomial

É possível especificar as seguintes opções para Regressão logística multinomial:

**Escala de dispersão.** Permite especificar o valor de escala de dispersão que será usado para corrigir a estimativa da matriz de covariância paramétrica. **Deviance** estima o valor de escala usando a estatística da função deviance (qui-quadrado de razão de verossimilhança). **Pearson** estima o valor de escala usando a estatística qui-quadrado de Pearson. Também é possível especificar seu próprio valor de escala. Ele deve ser um valor numérico positivo.

**Opções stepwise.** Essas opções fornecem controle dos critérios estatísticos quando métodos stepwise são usados para construir um modelo. Elas são ignoradas, a menos que um modelo stepwise seja especificado na caixa de diálogo Modelo.

- **Probabilidade de entrada.** Essa é a probabilidade da estatística de razão de verossimilhança para a entrada de variável. Quando maior a probabilidade especificada, mais fácil será para uma variável entrar no modelo. Esse critério será ignorado, a menos que o método de entrada forward, stepwise forward ou stepwise backward seja selecionado.
- **Teste de entrada.** Este é o método para inserir termos em métodos stepwise. Escolha entre o teste de razão de verossimilhança e o teste de escore. Esse critério será ignorado, a menos que o método de entrada forward, stepwise forward ou stepwise backward seja selecionado.
- **Probabilidade de remoção.** Essa é a probabilidade da estatística da razão de verossimilhança para a remoção de variável. Quando maior a probabilidade especificada, mais fácil será para uma variável permanecer no modelo. Esse critério será ignorado, a menos que o método de eliminação backward, stepwise forward ou stepwise backward seja selecionado.
- **Teste de remoção.** Este é o método para remover termos em métodos stepwise. Escolha entre o teste de razão de verossimilhança e o teste de Wald. Esse critério será ignorado, a menos que o método de eliminação backward, stepwise forward ou stepwise backward seja selecionado.
- **Mínimo de efeitos graduais no modelo.** Ao usar os métodos de eliminação backward ou stepwise backward, isso especifica o número mínimo de termos a serem incluídos no modelo. O intercepto não é contado como um termo modelo.
- **Máximo de efeitos graduais no modelo.** Ao usar os métodos de entrada forward ou stepwise forward, isso especifica o número máximo de termos a serem incluídos no modelo. O intercepto não é contado como um termo modelo.
- **Restringir hierarquicamente a entrada e remoção de termos.** Essa opção permite escolher se colocar restrições na inclusão de termos modelo. A hierarquia requer que, para qualquer termo a ser incluído, todos os termos de ordem inferior que fazem parte do termo a ser incluído devem estar no modelo primeiro. Por exemplo, se o requisito de hierarquia estiver em vigor, os fatores *Estado civil* e *Gênero* devem estar no modelo antes da inclusão da interação de *Estado civil\*Gênero*. As três opções de botões de opções determinam a função de covariáveis na determinação da hierarquia.

## Regressão logística multinomial - Salvar

A caixa de diálogo Salvar permite salvar variáveis no arquivo de trabalho e exportar informações de modelo para um arquivo externo.

**Variáveis salvas.** As seguintes variáveis podem ser salvas:

- **Probabilidades de resposta estimada.** Essas são as probabilidades estimadas de classificação de um padrão de fator/covariável nas categorias de resposta. Existem tantas probabilidades estimadas quanto categorias da variável de resposta; até 25 serão salvas.
- **Categoria predita.** Essa é a categoria de resposta com a maior probabilidade esperada para um padrão de covariável/fator.
- **Probabilidades de categoria predita.** Este é o máximo das probabilidades de resposta estimadas.
- **Probabilidade de categoria real.** Essa é a probabilidade estimada de classificação de um padrão de covariável/fator na categoria observada.

**Exportar informações de modelo em arquivo XML.** As estimativas paramétrica e (opcionalmente) suas covariâncias são exportadas no arquivo especificado em formato XML (PMML). É possível usar esse arquivo de modelo para aplicar as informações de modelo a outros arquivos de dados para propósitos de escoragem.

## Recursos adicionais do comando NOMREG

O idioma da sintaxe de comando também permite:

- Especificar a categoria de referência da variável dependente.
- Incluir casos com valores omissos de usuário.
- Customizar testes de hipótese especificando hipóteses nulas como combinações lineares paramétricas.

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

## Regressão probito

---

Esse procedimento mede o relacionamento entre a intensidade de um estímulo e a proporção de casos exibindo uma determinada resposta ao estímulo. Ele é útil para situações nas quais você tem uma saída dicotômica que é considerada como influenciada ou causada por níveis de algumas variáveis independentes e é bem apropriado principalmente para dados experimentais. Esse procedimento permitirá estimar a intensidade de um estímulo necessário para induzir a uma determinada proporção de respostas, como a dose efetiva mediana.

**Exemplo.** Qual é a eficiência de um novo pesticida em matar formigas e qual a concentração apropriada a ser usada? Você pode fazer uma experiência na qual expõe amostras de formigas a diferentes concentrações do pesticida e, em seguida, registra o número de formigas mortas e o número de formigas expostas. Ao aplicar a regressão probito a esses dados, será possível determinar a intensidade do relacionamento entre a concentração e o extermínio. Também, será possível determinar qual a concentração apropriada de pesticida, se você quiser ter certeza de matar, digamos, 95% das formigas expostas.

**Estatísticas.** Coeficientes de regressão e erros padrão, interceptação e erro padrão, qui-quadrado da qualidade do ajuste de Pearson, frequências observadas e esperadas e intervalos de confiança para níveis efetivos de variáveis independentes. Gráficos: Gráficos de respostas transformados.

Este procedimento utiliza os algoritmos propostos e implementados no NPSOL<sup>®</sup> por Gill, Murray, Saunders e Wright, para estimar os parâmetros do modelo.

### Considerações de dados da regressão probito

**Dados.** Para cada valor da variável independente (ou cada combinação de valores para diversas variáveis independentes), sua variável de resposta deve ser uma contagem do número de casos com esses valores que mostram a resposta de interesse, e a variável observada total deve ser uma contagem do número total de casos com esses valores para a variável independente. A variável de fator deve ser categórica, codificada como números inteiros.

**Suposições.** As observações devem ser independentes. Se você tiver um grande número de valores para as variáveis independentes relativas ao número de observações, como é possível em um estudo observacional, as estatísticas de qui-quadrado e de qualidade de ajuste podem não ser válidas.

**Procedimentos relacionados.** A análise de probit está estritamente relacionada à regressão logística; de fato, se você escolher a transformação logit, esse procedimento basicamente calculará uma regressão logística. Em geral, a análise de probit é apropriada para experiências projetadas, enquanto a regressão logística é mais apropriada para estudos observacionais. As diferenças na saída refletem essas diferentes ênfases. O procedimento de análise de probit relata estimativas de valores efetivos para várias taxas de resposta (incluindo a dose efetiva mediana), enquanto o procedimento de regressão logística relata estimativas de razões de chances para variáveis independentes.

### Obtendo uma análise da regressão probito

1. Nos menus, escolha:

**Analisar > Regressão > Probit...**

2. Selecione uma variável de frequência de resposta. Essa variável indica o número de casos exibindo uma resposta ao estímulo de teste. Os valores dessa variável não podem ser negativos.

3. Selecione uma variável observada total. Essa variável indica o número de casos aos quais o estímulo foi aplicado. Os valores dessa variável não podem ser negativos e não podem ser menores que os valores da variável de frequência de resposta para cada caso.

Opcionalmente, é possível selecionar uma variável de Fator. Se você o fizer, use **Definir Intervalo** para definir o intervalo para os grupos.

4. Selecione uma ou mais covariáveis. Esta variável contém o nível do estímulo aplicado a cada observação. Se desejar transformar a covariável, selecione uma transformação na lista suspensa **Transformação**. Se nenhuma transformação for aplicada e houver um grupo de controle, o grupo de controle será incluído na análise.
5. Selecione o modelo **Probit** ou **Logit**.

#### **Modelo Probit**

Aplica a transformação probito (o inverso da função de distribuição normal padrão acumulativa) às proporções de resposta.

#### **Modelo Logit**

Aplica-se a transformação logit (chances de log) às proporções de resposta.

## **Definir intervalo de Análise de probito**

Isso permite especificar os níveis da variável fator que serão analisados. Os níveis do fator devem ser codificados como números inteiros consecutivos e todos os níveis no intervalo especificado serão analisados.

## **Opções de análise de probit**

É possível especificar opções para sua análise de probit:

**Estatísticas.** Permite solicitar as seguintes estatísticas opcionais: Frequências, Potência média relativa, Teste de paralelismo e Intervalos de confiança fiduciários.

- *Potência Média Relativa.* Exibe a razão de potências medianas para cada par de níveis de fator. Além disso, mostra os limites de confiança de 95% para cada potência média relativa. As potências médias relativas não estarão disponíveis se você não tiver uma variável de fator ou se você tiver mais de uma covariável.
- *Teste de Paralelismo.* Um teste da hipótese de que todos os níveis do fator têm uma inclinação comum.
- *Intervalos de Confiança Fiduciários.* Intervalos de confiança para a dosagem do agente necessária para produzir uma certa probabilidade de resposta.

Os intervalos de confiança fiduciários e a Potência média relativa estarão indisponíveis se você tiver selecionado mais de uma covariável. A Potência média relativa e o Teste de paralelismo estarão disponíveis apenas se você tiver selecionado uma variável de fator.

**Taxa de resposta natural.** Permite indicar uma taxa de resposta natural mesmo na ausência do estímulo. As alternativas disponíveis são Nenhum, Calcular a partir de dados ou Valor.

- *Calcular a partir dos Dados.* Estimar a taxa de resposta natural dos dados de amostra. Seus dados devem conter um caso representando o nível de controle, para o qual o valor das covariáveis é de 0. O probito estima a taxa de resposta natural utilizando a proporção de respostas para o nível de controle como um valor inicial.
- *Valor.* Configura a taxa de resposta natural no modelo (selecione esse item quando souber a taxa de resposta natural antecipadamente). Insira a proporção de resposta natural (a proporção deve ser menor que 1). Por exemplo, se a resposta ocorrer 10% do tempo quando o estímulo for 0, digite 0,10.

**Crítérios.** Permite controlar parâmetros do algoritmo de estimação paramétrica iterativo. É possível substituir os padrões para Máximo de iterações, Limite de passos e Tolerância de otimalidade.

## **Recursos adicionais do comando PROBIT**

O idioma da sintaxe de comando também permite:

- Solicitar uma análise nos modelos probit e logit.
- Controlar o tratamento de valores omissos.
- Transformar as covariáveis por bases em vez de base 10 ou logaritmo natural.

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

## Regressão de quantil

Regressão é um método estatístico amplamente usado em modelagem quantitativa. A regressão linear múltipla é uma abordagem básica e padrão na qual os pesquisadores usam os valores de diversas variáveis para explicar ou prever os valores médios de um resultado de escala. No entanto, em muitas circunstâncias, estamos mais interessados no mediano ou em um quantil arbitrário do resultado de escala.

A regressão de quantil modela o relacionamento entre um conjunto de variáveis preditoras (independentes) e percentis específicos (ou "quantis") de uma variável de destino (dependente), mais frequentemente a mediana. Ela tem duas principais vantagens sobre a regressão por quadrados mínimos comum:

- A regressão de quantil não faz suposições sobre a distribuição da variável de destino.
- A regressão de quantil tende a resistir à influência de observações remotas

A regressão de quantil é amplamente usada para pesquisas em segmentos de mercado, como ecologia, assistência médica e economia financeira.

### Exemplo:

Qual é o relacionamento entre a renda doméstica total e a proporção de renda que é gasta com alimentação? A lei de Engels é uma observação na área de economia que afirma que, conforme a renda aumenta, a proporção de renda gasta com alimentação cai, mesmo que o gasto absoluto com alimentação aumente. Ao aplicar a regressão de quantil a esses dados, é possível determinar qual gasto com alimentação pode cobrir 90% das famílias (para 100 famílias com uma determinada renda) quando não estão interessadas na despesa média com alimentação.

### Estatísticas

Regressão de quantil, abordagem simplex, algoritmo de otimização não linear de pontos interiores de Frisch-Newton, Barrodale e Roberts, Bofinger, Hall Sheather, largura da banda, nível de significância, manipulações de matrizes, critério de convergência, ponderações de regressão, termo de intercepto, destino predito, resíduos de predição, tabulação, gráficos de predição, estimativas de parâmetros, matriz de covariâncias, matriz de correlações, valores observados, intervalo de confiança.

Esse procedimento usa os algoritmos propostos por Koenker, R. W. e Bassett, G. W. (1978). *Regression quantiles*, *Econometrica*, 46, 33–50.

## Considerações sobre dados de regressão de quantil

### Dados

É necessária uma variável dependente numérica única. A variável de destino precisa ser uma variável contínua. Os preditores podem ser variáveis contínuas ou variáveis dummy para preditores categóricos. O termo de intercepto ou pelo menos um preditor é necessário para executar uma análise.

### Suposições

A regressão de quantil não faz suposições sobre a distribuição da variável de destino e resiste à influência de observações remotas.

### Procedimentos Relacionados

A análise de quantil está relacionada à regressão por quadrados mínimos ordinários.

## Obtendo uma análise de regressão de quantil

1. Nos menus, escolha:



## Analisar > Regressão > Quantil...

O diálogo permite especificar as variáveis de destino, fator, covariável e de ponderação a serem usadas para análise de regressão de quantil. O diálogo também fornece a opção de conservar memória para análise complexa ou grandes conjuntos de dados.

2. Selecione uma variável de destino numérica. Somente uma variável de destino é necessária para executar uma análise. Somente variáveis numéricas são permitidas.
3. Opcionalmente, selecione uma ou mais variáveis fator. Variáveis de escala não são permitidas.
4. Opcionalmente, selecione uma ou mais variáveis covariáveis. Variáveis de sequência não são permitidas.

**Nota:** Quando as listas **Fator(es)** e **Covariável(is)** estiverem vazias e **Incluir intercepto no modelo** for selecionado no diálogo Modelo, a seguinte mensagem será exibida:

No effects have been specified. Therefore, an intercept only model will be fit.  
Do you want to fit an intercept-only model?

5. Opcionalmente, selecione uma variável de ponderação de regressão. Variáveis de sequência não são permitidas.
6. Opcionalmente, selecione **Conservar memória para análise complexa ou grandes conjuntos de dados**. Esta configuração controla se os dados ficarão ou não retidos em um arquivo externo durante o processamento. A ativação da configuração pode ajudar a conservar recursos de memória ao executar análises complexas ou análises com grandes conjuntos de dados.

## Regressão de quantil: Critérios

O diálogo Critérios fornece opções para

### Quantil

Fornece opções para especificar os quantis.

#### Especificar quantis únicos

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

Todos os valores devem ser exclusivos (não são permitidos valores duplicados). O valor padrão é 0.5.

#### Especificar quantis de grade

Quando selecionada, uma grade de quantis pode ser especificada de um valor **inicial** (value1) a um valor **final** (value2) com o incremento de **By** (value3). Se especificado, somente um conjunto válido de  $[value1 \text{ TO } value2 \text{ BY } value3]$  é permitido. Ele deve satisfazer esse  $0 \leq value1 \leq value2 \leq 1$ . Nos casos em que  $value1 = value2$ , isso é equivalente a especificar um único value1, independentemente de value3.

### Método de estimação

Fornece opções para especificar o método de estimação de modelo.

#### Escolhido automaticamente pelo programa

Permite que o procedimento selecione automaticamente o método de estimação apropriado. Esta é a definição padrão.

#### Algoritmo simplex

Chama o algoritmo simplex que foi desenvolvido por Barrodale e Roberts.

#### Otimização não linear de pontos interiores de Frisch-Newton

Chama o algoritmo de otimização não linear de pontos interiores de Frisch-Newton.

### Pós-estimação

Fornece opções para a pós-estimação da variância-covariância das estimativas de parâmetros e dos intervalos de confiança para os valores de destino preditos.

**Considerar que os casos são IID**

Quando selecionada, esta configuração considera que os termos de erro são distribuídos de forma independente e idêntica. Quando a configuração não é selecionada, o tempo de cálculo pode aumentar significativamente para modelos grandes. A configuração é selecionada por padrão.

**Tipo de largura da banda**

Determina qual método de largura da banda é usado para estimar a matriz de variância-covariância das estimativas de parâmetros (**Bofinger** ou **Hall-Sheather**). **Bofinger** é a configuração padrão.

**Método Numérico**

Fornece as seguintes opções:

**Tolerância de singularidade**

Especifica o valor de tolerância para as manipulações de matrizes no método de ponto interior. O valor especificado deve ser um valor duplo único em  $(0, 10^{-3})$ , com  $10^{-12}$  como a configuração padrão.

**Convergência**

Especifica o critério de convergência para o método numérico. O valor especificado deve ser um valor duplo único em  $(0, 10^{-3})$ , com  $10^{-6}$  como a configuração padrão.

**Máximo de iterações**

Especifica o número máximo de iterações. O valor especificado deve ser um número inteiro positivo único. O valor padrão é 2000.

**Valores omissos**

Fornece opções para determinar como os valores omissos são manipulados.

**Excluir valores omissos do usuário e do sistema**

Quando selecionada, os valores omissos do usuário e do sistema são excluídos.

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

Quando selecionada, os valores omissos do usuário são tratados como válidos.

**Intervalo de confiança (%)**

Especifica o nível de significância. Quando especificada, o valor deve ser um valor duplo único entre 0 e 100. O valor padrão é 95.

## Regressão de quantil: Modelo

O diálogo Modelo fornece opções para especificar os efeitos e as ponderações que são usados no modelo. Se omitido ou especificado por si próprio, o modelo conterá o termo de intercepto e todos os efeitos principais com as covariáveis na lista de covariáveis e os fatores na lista de fatores.

**Especificar efeitos do modelo**

O modelo padrão é somente de interceptação; portanto, deve-se especificar explicitamente outros efeitos do modelo. Como alternativa, é possível criar termos aninhados e não aninhados. Quando **Construir termos** é selecionado, as seguintes opções de efeito e interação estão disponíveis para termos não agrupados.

**Efeitos principais**

Cria um termo dos principais efeitos para cada variável selecionada.

**Interação**

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

**Fatorial**

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

**Todos de 2 fatores**

Cria todas as interação de dois fatores das variáveis selecionadas.

**Todos de 3 fatores**

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

**Todos de 4 fatores**

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

### Todos de 5 fatores

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

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

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

### Notas sobre agrupamento:

- Para incluir um efeito para uma interação entre dois fatores, use a palavra-chave BY ou o asterisco (\*) para juntar os fatores que estão envolvidos na interação.
- Os fatores dentro de um efeito de interação devem ser distintos.
- Use pares de parênteses para incluir um efeito para agrupar um termo dentro de outro.
- Quando mais de um par de parênteses estiver presente, cada par de parênteses deve ser colocado ou agrupado dentro de outro par de parênteses.
- O agrupamento múltiplo é permitido.
- Interações entre efeitos agrupados não são suportadas.

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

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

### Incluir intercepto no modelo

Quando selecionada, o termo de intercepto é incluído no modelo. Quando não selecionada, pelo menos um preditor é necessário para executar a análise. A configuração é ativada por padrão.

## Regressão de quantil: Exibição

O diálogo Exibição fornece configurações de controle de saída e de gráfico.

### Imprimir

As seguintes opções de saída estão disponíveis.

#### Estimativas de parâmetro

Exibe estimativas de parâmetro, estatísticas do teste correspondentes e intervalos de confiança.

É possível opcionalmente exibir estimativas de parâmetros exponenciado além das estimativas de parâmetros bruto.

#### Matriz de covariâncias para estimativas de parâmetro

Exibe a matriz de covariâncias paramétrica estimado.

#### Matriz de correlações para estimativas de parâmetro

Exibe a matriz de correlações paramétrica estimado.

### Plotar e tabular

As seguintes opções plotagem estão disponíveis:

#### Plotar estimativas de parâmetros para

É possível selecionar para plotar estimativas de parâmetros para um número especificado de efeitos principais, ou para todos os efeitos no modelo. A configuração **Principais xx efeitos** controla o número de categorias ou combinações de categorias em um efeito misto, que são

plotadas na interação de uma covariável e um ou dois fatores. O valor deve ser um número inteiro positivo único (50 é a configuração padrão).

**Notas:**

- Os gráficos de predição são criados para todos os efeitos quando o valor de número inteiro especificado é maior que o número de categorias ou combinações.
- A configuração é válida somente quando diversos valores são especificados para a configuração **Valores de quantil** no diálogo Critérios. Nenhum gráfico é criado quando é especificado um valor de quantil único.

**Exibir o gráfico de preditos versus observados**

Controla a criação do gráfico de valores preditos versus observados. Quando ativada, é criado um gráfico único que contém os pontos (com diferentes pontos coloridos representando diferentes quantis). A configuração é desativada por padrão.

**Prever os efeitos no modelo**

Quando ativada, as seguintes opções estão disponíveis:

**Plotar ou tabular os principais x efeitos**

Especifique o número de principais efeitos cujo gráfico de predição ou tabela de predição será criada. 3 é o valor padrão.

**Nota:** Os gráficos de predição ou tabelas de predição são criados para todos os efeitos quando o valor especificado é maior que o número de efeitos válidos no modelo.

**Plotar ou tabular efeitos especificados pelo usuário**

As diretrizes de efeitos válidas são:

- Efeito com uma covariável (incluindo uma alta potência da própria covariável): Crie um único gráfico contendo as linhas preditas por diferentes quantis.
- Efeito com um fator: Tabule as predições para as categorias do fator por diferentes quantis.
- Efeito com a interação de dois fatores: Para cada quantil, tabule as predições para as categorias de dois fatores.
- Efeito com a interação de uma covariável e um ou dois fatores: Para cada quantil, crie um gráfico contendo as linhas para cada categoria ou combinação das categorias no efeito de interação.
- O número máximo de combinações a serem plotadas é controlado pelo valor especificado para **Plotar um máximo de xx categorias de combinações de categorias em um efeito misto**.

Os efeitos que são movidos de **Efeitos de modelo** para a lista **Linhas de predição** são usados para plotagem. Os gráficos não são criados nos casos em que os efeitos especificados são constantes (removidos da construção de modelo).

Os efeitos que são movidos de **Efeitos de modelo** para a lista **Tabelas de predição** são usados para tabulação. As tabelas não são criadas nos casos em que os efeitos especificados são constantes (removidos da construção de modelo).

**Plote um máximo de xx categorias ou combinações de categorias em um efeito misto**

Controla o número máximo das combinações de categorias a serem plotadas. O valor padrão é 10.

## Regressão de quantil: Salvar

O diálogo Salvar fornece opções para escoragem do modelo.

**Valor predito de resposta**

Quando selecionada, o valor de destino predito é escorado.

**Resíduo**

Quando selecionada, os resíduos de predição são escorados.

### Limite inferior de intervalo de predição

Quando selecionada, os limites inferiores dos intervalos de predição são escorados.

### Limite superior de intervalo de predição

Quando selecionada, os limites superiores dos intervalos de predição são escorados.

**Nota:** Um nome de variável pode ser especificado para cada opção de salvamento. Se um nome raiz for especificado, ele deverá ser um nome de variável válido. Um nome raiz, seguido por um caractere de sublinhado "\_" e um sufixo de quantil significativo, é usado quando diversos valores são especificados para a configuração **Valores de quantil** no diálogo Critérios.

## Regressão de quantil: Exportação

O diálogo Exportação fornece opções para especificar quais estatísticas são exportadas, como as estatísticas são exportadas (arquivo de dados ou conjuntos de dados externos) e controlar como os dados são manipulados durante o processamento (processados normalmente ou retidos em um arquivo temporário durante o processamento).

### Matriz de covariâncias de estimativas de parâmetro

Quando selecionada, as opções para gravar a matriz de covariâncias das estimativas de parâmetros em um arquivo de dados externo ou em um conjunto de dados declarado anteriormente são ativadas.

### Matriz de correlações de estimativas de parâmetro

Quando selecionada, as opções para gravar a matriz de correlações das estimativas de parâmetros em um arquivo de dados externo ou em um conjunto de dados declarado anteriormente são ativadas.

### A matriz de covariâncias/correlações será salva em um conjunto de dados único ou arquivo externo na presença de vários quantis de regressão

Quando vários quantis estão presentes, esta opção alterna o salvamento de matrizes de covariâncias/correlações para conjuntos de dados únicos ou múltiplos ou arquivos de dados externos. Quando não ativada, as matrizes são salvas em um arquivo de dados ou conjunto de dados único e externo. Quando ativada, as matrizes são salvas em múltiplos arquivos de dados ou conjuntos de dados externos. A configuração entra em vigor somente quando diversos valores são especificados para a configuração **Valores de quantil** no diálogo Critérios.

**Nota:** Esta opção está disponível somente quando **Matriz de covariâncias de estimativas de parâmetros** ou **Matriz de correlações de estimativas de parâmetros** é selecionada.

### Exportar informações do modelo para o arquivo XML

Quando selecionada, fornece opções para exportar as informações de modelo para um nome e local de arquivo XML específico.

### Exportar como XML

Quando **Exportar informações de modelo para arquivo XML** é selecionado, é possível selecionar para exportar estimativas de parâmetros e matrizes de covariâncias ou somente estimativas de parâmetros. **Estimativas de parâmetros e matriz de covariâncias** é a configuração padrão.

## Convenções de nome do arquivo

- Quando um valor único é especificado para a configuração **Valores de quantil** no diálogo Critérios, `savefile` e `dataset` são usados para nomear o arquivo de dados ou conjunto de dados externo.
- Quando diversos valores são especificados para a configuração **Valores de quantil** no diálogo Critérios, cada quantil é salvo em um arquivo de dados ou conjunto de dados externo.
- Um caractere de sublinhado "\_", seguido por um sufixo de quantil significativo, é anexado automaticamente ao nome do arquivo de dados ou do conjunto de dados. Por exemplo, quando 0,25, 0,50 e 0,75 são especificados como **Valores de quantil**, os sufixos `_25`, `_50` e `_75` são anexados aos nomes de arquivos de dados (antes da extensão `.sav`).
- Dígitos adicionais podem ser especificados para cada sufixo de quantil (se necessário).
- O zero à esquerda e a vírgula decimal de **Valores de quantil** não são usados no sufixo.
- Quando a notação científica é especificada para **Valores de quantil**, ela é convertida em um valor decimal quando exibida no sufixo.

## Regressão não linear

A regressão não linear é um método de localizar um modelo não linear do relacionamento entre a variável dependente e um conjunto de variáveis independentes. Diferentemente da regressão linear tradicional, que é restrita à estimativa de modelos lineares, a regressão não linear pode estimar modelos com relacionamentos arbitrários entre variáveis independentes e dependentes. Isso é feito usando algoritmos de estimação iterativos. Observe que esse procedimento não é necessário para modelos polinomiais simples do formato  $Y = A + BX^2$ . Ao definir  $W = X^2$ , obtemos um modelo linear simples,  $Y = A + BW$ , que pode ser estimado usando métodos tradicionais, como o procedimento Regressão linear.

**Exemplo.** A população pode ser prevista com base no tempo? Um gráfico de dispersão mostra que parece haver um forte relacionamento entre população e tempo, mas o relacionamento é não linear, portanto, ele requer os métodos de estimação especiais do procedimento Regressão não linear. Ao configurar uma equação apropriada, como um modelo logístico de crescimento de população, podemos obter uma boa estimativa do modelo, permitindo fazer previsões sobre a população para os momentos que não foram realmente medidos.

**Estatísticas.** Para cada iteração: estimativas de parâmetros e soma de quadrados dos resíduos. Para cada modelo: soma de quadrados para regressão, resíduos, total não corrigido e total corrigido, estimativas de parâmetros, erros padrão assintóticos e matriz de correlação assintótica de estimativas de parâmetros.

**Nota:** A regressão não linear restrita usa os algoritmos propostos e implementados em NPSOL® por Gill, Murray, Saunders e Wright para estimar os parâmetros do modelo.

### Considerações de dados de regressão não linear

**Dados.** As variáveis dependentes e independentes devem ser quantitativas. Variáveis categóricas, como religião, maioridade ou região de residência, precisam ser recodificadas para variáveis binárias (fictícias) ou outros tipos de variáveis de contraste.

**Suposições.** Os resultados serão válidos apenas se você tiver especificado uma função que descreve com exatidão o relacionamento entre variáveis dependentes e independentes. Além disso, a escolha de um bons valores iniciais é muito importante. Mesmo que tenha especificado o formato funcional correto do modelo, se você usar valores iniciais não adequados, seu modelo poderá falhar ao convergir ou é possível obter uma solução ideal local em vez de uma ideal global.

**Procedimentos relacionados.** Muitos modelos que parecem ser não lineares inicialmente podem ser transformados em um modelo linear, que pode ser analisado usando o procedimento Regressão linear. Se você não tiver certeza de como deve ser o modelo apropriado, o procedimento Curva de estimação pode ajudar a identificar relações funcionais úteis em seus dados.

### Obtendo uma análise de regressão não linear

1. Nos menus, escolha:

**Analisar > Regressão > Não linear...**

2. Selecione uma variável dependente numérica da lista de variáveis em seu conjunto de dados ativo.
3. Para construir uma expressão de modelo, insira a expressão no campo **Modelar expressão** ou cole componentes (variáveis, parâmetros, funções) no campo.
4. Identifique parâmetros em seu modelo clicando em **Parâmetros**.

Um modelo segmentado (um que usa diferentes formatos em diferentes partes de seu domínio) deve ser especificado usando a lógica condicional na instrução de modelo único.

### Lógica condicional (Regressão não linear)

É possível especificar um modelo segmentado usando a lógica condicional. Para usar a lógica condicional em uma expressão de modelo ou em uma função de perda, forme a soma de uma série de termos, um

para cada condição. Cada termo consiste em uma expressão lógica (entre parênteses) multiplicada pela expressão que deve resultar quando essa expressão lógica for verdadeira.

Por exemplo, considere um modelo segmentado que equivale a 0 para  $X \leq 0$ ,  $X$  para  $0 < X < 1$ , and 1 for  $X \geq 1$ . A expressão para isto é:

$$(X \leq 0) * 0 + (X > 0 \ \& \ X < 1) * X + (X \geq 1) * 1.$$

Todas as expressões lógicas entre parênteses são avaliadas como 1 (true) ou 0 (false). Portanto:

Se  $X \leq 0$ , a expressão acima será reduzida para  $1 * 0 + 0 * X + 0 * 1 = 0$ .

Se  $0 < X < 1$ , ela será reduzida para  $0 * 0 + 1 * X + 0 * 1 = X$ .

Se  $X \geq 1$ , ela será reduzida para  $0 * 0 + 0 * X + 1 * 1 = 1$ .

Exemplos mais complicados podem ser facilmente construídos, substituindo diferentes expressões lógicas e expressões de resultado. Lembre-se das desigualdades duplas, como  $0 < X < 1$ , must be written as compound expressions, such as  $(X > 0 \ \& \ X < 1)$ .

As variáveis de sequência de caracteres podem ser usadas em expressões lógicas:

$$(\text{city} = \text{'New York'}) * \text{costliv} + (\text{city} = \text{'Des Moines'}) * 0.59 * \text{costliv}$$

Isso gera uma expressão (o valor da variável *costliv*) para nova-iorquinos e outro (59% desse valor) para residentes de Des Moines. As constantes da sequência devem ser colocadas entre aspas ou apóstrofes, conforme mostrado aqui.

## Parâmetros de regressão não linear

Parâmetros são partes de seu modelo estimadas pelo procedimento Regressão não linear. Os parâmetros podem ser constantes aditivas, coeficientes multiplicativos, expoentes ou valores usados na avaliação de funções. Todos os parâmetros que foram definidos aparecerão (com seus valores iniciais) na lista Parâmetros na caixa de diálogo principal.

**Nome.** Deve-se especificar um nome para cada parâmetro. Esse nome deve ser um nome de variável válido e deve ser o nome usado na expressão do modelo na caixa de diálogo principal.

**Valor inicial.** Permite especificar um valor inicial para o parâmetro, preferencialmente o mais próximo possível da solução final esperada. Valores de iniciais simples podem resultar em falha ao convergir ou na convergência em uma solução que é local (em vez de global) ou que é fisicamente impossível.

**Usar valores iniciais de análise anterior.** Se você já executou uma regressão não linear nessa caixa de diálogo, será possível selecionar essa opção para obter os valores iniciais paramétricas de seus valores na execução anterior. Isso permite continuar a procura quando o algoritmo está convergindo lentamente. (Os valores iniciais ainda aparecerão na lista Parâmetros na caixa de diálogo principal.)

*Nota:* Essa seleção persiste nessa caixa de diálogo durante o resto de sua sessão. Se você mudar o modelo, certifique-se de cancelar sua seleção.

## Modelos comuns de regressão não linear

A tabela a seguir fornece a sintaxe do modelo de exemplo para muitos modelos de regressão não linear publicados. Um modelo selecionado aleatoriamente provavelmente não ajustará bem os seus dados. Os valores iniciais apropriados para os parâmetros são necessários, e alguns modelos requerem restrições para convergir.

*Tabela 1. Sintaxe do modelo de exemplo*

Nome	Expressão do modelo
Regressão assintótica	$b1 + b2 * \exp(b3 * x)$
Regressão assintótica	$b1 - (b2 * (b3 ** x))$
Densidade	$(b1 + b2 * x) ** (-1 / b3)$

Tabela 1. Sintaxe do modelo de exemplo (continuação)

Nome	Expressão do modelo
Gauss	$b1 * (1 - b3 * \exp(-b2 * x^{**} 2))$
Gompertz	$b1 * \exp(-b2 * \exp(-b3 * x))$
Johnson-Schumacher	$b1 * \exp(-b2 / (x + b3))$
Modificado por log	$(b1 + b3 * x)^{**} b2$
Logística de log	$b1 - \ln(1 + b2 * \exp(-b3 * x))$
Lei dos Retornos Decrescentes de Metcherlich	$b1 + b2 * \exp(-b3 * x)$
Michaelis Menten	$b1 * x / (x + b2)$
Morgan-Mercer-Florin	$(b1 * b2 + b3 * x^{**} b4) / (b2 + x^{**} b4)$
Peal-Reed	$b1 / (1 + b2 * \exp(-(b3 * x + b4 * x^{**} 2 + b5 * x^{**} 3)))$
Razão de cúbicos	$(b1 + b2 * x + b3 * x^{**} 2 + b4 * x^{**} 3) / (b5 * x^{**} 3)$
Razão de quadráticos	$(b1 + b2 * x + b3 * x^{**} 2) / (b4 * x^{**} 2)$
Richards	$b1 / ((1 + b3 * \exp(-b2 * x))^{**} (1 / b4))$
Verhulst	$b1 / (1 + b3 * \exp(-b2 * x))$
Von Bertalanffy	$(b1^{**} (1 - b4) - b2 * \exp(-b3 * x))^{**} (1 / (1 - b4))$
Weibull	$b1 - b2 * \exp(-b3 * x^{**} b4)$
Densidade de lucro	$(b1 + b2 * x + b3 * x^{**} 2)^{**} (-1)$

## Função de perda de regressão não linear

A **função de perda** em regressão não linear é a função que é minimizada pelo algoritmo. Selecione **Soma de resíduos quadrados** para minimizar a soma dos resíduos quadrados ou **Função de perda definida pelo usuário** para minimizar uma função diferente.

Se você selecionar **Função de perda definida pelo usuário**, deverá definir a função de perda cuja soma (em todos os casos) deve ser minimizada pela opção de valores paramétrica.

- A maioria das funções de perda envolve a variável especial *RESID\_*, que representa o resíduo. (A Soma padrão de função de perda de resíduos quadrados pode ser inserida explicitamente como *RESID\_\*\*2*.) Se precisar usar o valor predito em sua função de perda, ele é igual à variável dependente menos o resíduo.
- É possível especificar uma função de perda condicional usando a lógica condicional.

É possível digitar uma expressão no campo Função de perda definida pelo usuário ou colar componentes da expressão no campo. As constantes de sequência devem ser colocadas entre aspas ou apóstrofes e as constantes numéricas devem ser digitadas em formato norte-americano, com o ponto como um delimitador decimal.

## Restrições paramétrica de regressão não linear

Uma **restrição** é uma restrição nos valores permitidos para um parâmetro durante a procura iterativa por uma solução. As expressões lineares são avaliadas antes de um passo ser executado, portanto, é possível usar restrições lineares para evitar passos que podem resultar em estouros. As expressões não lineares não avaliadas após a execução de um passo.

Cada equação ou desigualdade requer os seguintes elementos:

- Uma expressão envolvendo pelo menos um parâmetro no modelo. Digite a expressão ou use o teclado numérico, que permite colar números, operadores ou parênteses na expressão. É possível digitar os



parâmetros necessários junto com o restante da expressão ou colar da lista Parâmetros à esquerda. Não é possível usar variáveis ordinárias em uma restrição.

- Um dos três operadores lógicos são  $\leq$ ,  $=$  ou  $\geq$ .
- Uma constante numérica, com a qual a expressão é comparada usando o operador lógico. Digite a constante. As constantes numéricas devem ser digitadas em formato americano, com o ponto como um delimitador decimal.

## Regressão não linear - Salvar novas variáveis

É possível salvar diversas novas variáveis em seu arquivo de dados ativo. As opções disponíveis são Resíduos, Valores preditos, Derivados e valores de Função de perda. Essas variáveis podem ser usadas em análises subsequentes para testar o ajuste do modelo ou para identificar casos de problemas.

- *Residuais*. Salva os resíduos com o nome de variável resid\_.
- *Valores previstos*. Salva valores preditos com o nome de variável pred\_.
- *Derivados*. Um derivativo é salvo para cada parâmetro de modelo. Os nomes derivados são criados ao prefixar 'd.' nos seis primeiros caracteres de nomes do parâmetro.
- *Valores de Função de Perda*. Esta opção está disponível se você especificar sua própria função de perda. O nome de variável \_loss é designado aos valores da função de perda.

## Opções de regressão não linear

As Opções permitem controlar vários aspectos de sua análise de regressão não linear:

*Estimativas bootstrap*. Um método de estimar o erro padrão de uma estatística usando amostras repetidas do conjunto de dados original. Isso é feito por amostragem (com substituição) para obter muitas amostras do mesmo tamanho que o conjunto de dados original. A equação não linear é estimada para cada uma destas amostras. Em seguida, o erro padrão de cada estimativa paramétrica é calculado como o desvio padrão das estimativas bootstrapped. Os valores dos parâmetros dos dados originais são utilizados como valores de início para cada amostra bootstrap. Isso requer o algoritmo de programação quadrática sequencial.

**Método de estimação**. Permite selecionar um método de estimação, se possível. (Algumas opções nessa ou em outras caixas de diálogo requerem o algoritmo de programação quadrática sequencial.) As alternativas disponíveis incluem programação quadrática sequencial e Levenberg-Marquardt.

- *Programação Quadrática Sequencial*. Este método está disponível para modelos restritos e não restritos. A programação quadrática sequencial será utilizada automaticamente se você especificar um modelo restrito, uma função de perda definida pelo usuário ou bootstrapping. É possível inserir novos valores para o Máximo de iterações e limite de Passo, bem como alterar a seleção nas listas suspensas para tolerância de Otimalidade, precisão de Função e o tamanho de passo Infinito.
- *Levenberg-Marquardt*. Este é o algoritmo padrão para modelos não restritos. O método de Levenberg-Marquardt não estará disponível se você especificar um modelo restrito, uma função de perda definida pelo usuário, ou bootstrapping. É possível inserir novos valores para o máximo de iterações e também alterar a seleção nas listas suspensas para convergência Soma de quadrados e convergência Parâmetro.

## Interpretando resultados de regressão não linear

Os problemas de regressão não linear geralmente apresentam dificuldades computacionais:

- A escolha de valores iniciais para os parâmetros influencia a convergência. Tente escolher valores iniciais que sejam razoáveis e, se possível, próximos da solução final esperada.
- Às vezes, um algoritmo tem melhor desempenho do que o outro em um determinado problema. No diálogo Opções, selecione o outro algoritmo se ele estiver disponível. (Se você especificar uma função de perda ou determinados tipos de restrições, não será possível usar o algoritmo de Levenberg-Marquardt.)

- Quando a iteração parar somente porque ocorreu o número máximo de iterações, o modelo "final" provavelmente não será uma boa solução. Selecione **Usar valores iniciais da análise prévia** no diálogo Parâmetros para continuar a iteração ou, melhor ainda, escolha valores iniciais diferentes.
- Os modelos que requerem exponenciação de ou por grandes valores de dados podem causar estouros ou estouros negativos (números muito grandes ou muito pequenos para o computador representar). Às vezes, é possível evitá-los por uma escolha adequada de valores iniciais ou impondo restrições nos parâmetros.

## Recursos adicionais do comando NLR

O idioma da sintaxe de comando também permite:

- Nomeie um arquivo a partir do qual ler valores iniciais para estimativas paramétrica.
- Especifique mais de uma instrução de modelo e função de perda. Isso facilita a especificação de um modelo segmentado.
- Forneça seus próprios derivados em vez de usar os calculados pelo programa.
- Especifique o número de amostras bootstrap a serem geradas.
- Especifique critérios de iteração adicionais, incluindo a configuração de um valor crítico para a verificação de derivados e a definição de um critério de convergência para a correlação entre os resíduos e os derivados.

Os critérios adicionais para o comando CNLR (regressão não linear restrita) permitem:

- Especificar o número máximo de iterações menores permitidas em cada iteração principal.
- Configurar um valor crítico para a verificação de derivados.
- Configurar um limite de passos.
- Especificar uma tolerância de travamento para determinar se os valores iniciais estão dentro de seus limites especificados.

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

## Estimação Ponderada

Os modelos de regressão linear padrão consideram que a variância é constante na população em estudo. Quando isso não for o caso (por exemplo, quando os casos que forem altos em algum atributo mostrarem mais variabilidade do que os casos que forem baixos nesse atributo) a regressão linear que usa quadrados mínimos ordinários (OLS) não fornecerá mais estimativas de modelo ideais. Se as diferenças em variabilidade puderem ser preditas a partir de outra variável, o procedimento Estimação ponderada poderá calcular os coeficientes de um modelo de regressão linear usando quadrados mínimos ponderados (WLS), de forma que as observações mais precisas (ou seja, as com menor variabilidade) recebem maior ponderação na determinação dos coeficientes de regressão. O procedimento Estimação ponderada testa um intervalo de transformações de ponderação e indica qual fornecerá o melhor ajuste para os dados.

**Exemplo.** Quais são os efeitos da inflação e do desemprego em mudanças de preços de ações? Como as ações com valores mais altos geralmente mostram maior variabilidade do que as de valores mais baixos, os quadrados mínimos ordinários não produzirão estimativas ideais. A estimação ponderada permite contar com o efeito do preço de ações na variabilidade de mudanças de preço no cálculo do modelo linear.

**Estatísticas.** Valores de log da verossimilhança para cada potência da variável de origem de ponderação testada,  $R$  múltiplo,  $R$ -quadrado,  $R$ -quadrado ajustado, tabela ANOVA para o modelo WLS, estimativas de parâmetros não padronizadas e padronizadas, e log da verossimilhança para o modelo WLS.

### Considerações de dados de estimação ponderada

**Dados.** As variáveis dependentes e independentes devem ser quantitativas. Variáveis categóricas, como religião, maioridade ou região de residência, precisam ser recodificadas para variáveis binárias (fictícias)

ou outros tipos de variáveis de contraste. A variável de ponderação deve ser quantitativa e deve estar relacionada à variabilidade na variável dependente.

**Suposições.** Para cada valor da variável independente, a distribuição da variável dependente deve ser normal. O relacionamento entre a variável dependente e cada variável independente deve ser linear, e todas as observações devem ser independentes. A variância da variável dependente pode variar nos níveis das variáveis independentes, mas as diferenças devem ser previsíveis com base na variável de ponderação.

**Procedimentos relacionados.** O procedimento Exploração pode ser usado para triar seus dados. A Exploração fornece testes para normalidade e homogeneidade de variância, bem como exibições gráficas. Se sua variável dependente parecer ter uma variância igual nos níveis de variáveis independentes, será possível usar o procedimento Regressão linear. Se seus dados parecerem violar uma suposição (como normalidade), tente transformá-los. Se seus dados não estiverem relacionados linearmente e uma transformação não ajudar, use um modelo alternativo no procedimento Curva de estimação. Se sua variável dependente for dicotômica (por exemplo, se uma determinada venda for concluída ou se um item estiver com defeito) use o procedimento Regressão logística. Se sua variável dependente for censurada (por exemplo, o tempo de sobrevivência após cirurgia) use Tabelas de mortalidade, Kaplan-Meier ou Regressão de Cox, disponíveis em Tabelas customizadas e Estatísticas avançadas. Se seus dados não forem independentes (por exemplo, se você observar a mesma pessoa em várias condições) use o procedimento Medidas repetidas, disponível em Tabelas customizadas e Estatísticas avançadas.

## Obtendo uma análise de estimação ponderada

1. Nos menus, escolha:

**Analisar > Regressão > Estimativa de ponderação...**

2. Selecione uma variável dependente.

3. Selecione uma ou mais variáveis independentes.

4. Selecione a variável que é a origem de heterocedasticidade como a variável de ponderação.

### **Variável de ponderação**

Os dados são ponderados pelo recíproco dessa variável elevada a uma potência. A equação de regressão é calculada para cada um dos valores de um intervalo de potência especificado e indica a potência que maximiza a função do log da verossimilhança.

### **Faixa de energia**

Isso é usado em conjunto com a variável de ponderação para calcular ponderações. Várias equações de regressão serão ajustadas, uma para cada valor no intervalo de potências. Os valores inseridos na caixa de teste de intervalo de Potência e na caixa de texto devem estar entre -6,5 e 7,5, inclusive. Os valores de potência variam do valor mais baixo para o mais alto, em incrementos determinados pelo valor especificado. O número total de valores no intervalo de potências é limitado a 150.

## Opções de estimação ponderada

É possível especificar opções para sua análise de estimação ponderada:

**Salvar melhor ponderação como nova variável.** Inclui a variável de ponderação no arquivo ativo. Essa variável é chamada *WGT\_n*, em que *n* é um número escolhido para dar um nome exclusivo à variável.

**Exibir ANOVA e estimativas.** Permite controlar como as estatísticas são exibidas na saída. As alternativas disponíveis são Para melhor potência e Para cada valor de potência.

## Recursos adicionais do comando WLS

O idioma da sintaxe de comando também permite:

- Fornecer um valor único para a potência.
- Especificar uma lista de valores de potência, ou misturar um intervalo de valores com uma lista de valores para a potência.

## Regressão por quadrados mínimos de dois estágios

Os modelos de regressão linear padrão consideram que os erros na variável dependente não estão correlacionados às variáveis independentes. Quando esse não é o caso (por exemplo, quando os relacionamentos entre variáveis são bidirecionais), a regressão linear usando quadrados mínimos ordinários (OLS) não fornece mais estimativas de modelo ideal. A regressão de quadrados mínimos de dois estágios usa variáveis instrumentais que não estão correlacionadas aos termos de erro para calcular valores estimados dos preditores problemáticos (o primeiro estágio) e, em seguida, usa esses valores calculados para estimar um modelo de regressão linear da variável dependente (o segundo estágio). Como os valores calculados são baseados em variáveis que não são correlacionadas aos erros, os resultados do modelo de dois estágios são os ideais.

**Exemplo.** A demanda por uma mercadoria está relacionada a seu preço e às receitas de consumidores? A dificuldade para este modelo é que o preço e a demanda têm um efeito recíproco entre si. Ou seja, o preço pode influenciar a demanda e a demanda também pode influenciar o preço. Um modelo de regressão de quadrados mínimos de dois estágios pode usar receitas de consumidores e preço defasado para calcular o preço de um proxy que não está correlacionado ao erros de medição na demanda. Esse proxy é substituído pelo próprio preço no modelo especificado originalmente, que é então estimado.

**Estatísticas.** Para cada modelo: coeficientes de regressão padronizados e não padronizados,  $R$  múltiplo,  $R^2$ ,  $R^2$  ajustado, erro padrão da estimativa, tabela de análise de variância, valores preditos e resíduos. Além disso, 95% de intervalos de confiança para cada coeficiente de regressão e matrizes de correlação e de covariâncias de estimativas de parâmetros.

### Considerações de dados de Regressão por quadrados mínimos de dois estágios

**Dados.** As variáveis dependentes e independentes devem ser quantitativas. Variáveis categóricas, como religião, maioridade ou região de residência, precisam ser recodificadas para variáveis binárias (fictícias) ou outros tipos de variáveis de contraste. As variáveis explicativas *endógenas* devem ser quantitativas (não categóricas).

**Suposições.** Para cada valor da variável independente, a distribuição da variável dependente deve ser normal. A variância da distribuição da variável dependente deve ser constante para todos os valores da variável independente. O relacionamento entre a variável dependente e cada variável independente deve ser linear.

**Procedimentos relacionados.** Se você acreditar que nenhuma de suas variáveis preditoras esteja correlacionada aos erros em sua variável dependente, será possível usar o procedimento Regressão linear. Se seus dados parecerem violar uma das suposições (como normalidade ou variância constante), tente transformá-los. Se seus dados não estiverem relacionados linearmente e uma transformação não ajudar, use um modelo alternativo no procedimento Curva de estimação. Se sua variável dependente for dicotômica, por exemplo, se uma determinada venda for ou não concluída, use o procedimento Regressão logística. Se os seus dados não são independentes--por exemplo, se você observar a mesma pessoa em várias condições--use o procedimento Medidas Repetidas.

### Obtendo uma análise de regressão de quadrados mínimos de dois estágios

1. Nos menus, escolha:

**Analisar > Regressão > 2-Estágio de quadrados mínimos...**

2. Selecione uma variável dependente.

3. Selecione uma ou mais variáveis explicativas (preditoras).

4. Selecione uma ou mais variáveis instrumentais.

- *Instrumental.* Estas são as variáveis usadas para calcular os valores preditos para as variáveis endógenas no primeiro estágio da análise de quadrados mínimos de dois estágios. As mesmas variáveis podem aparecer nas caixas de listagem Explicatória e Instrumental. O número de variáveis

instrumentais deve ser pelo menos igual ao número de variáveis explicatórias. Se todas as variáveis explicatórias e instrumentais listadas forem as mesmas, os resultados serão os mesmos que os resultados do procedimento de Regressão Linear.

As variáveis explicativas não especificadas como instrumentais são consideradas endógenas. Normalmente, todas as variáveis exógenas na lista Explicativa também são especificadas como variáveis instrumentais.

## Opções de regressão de quadrados mínimos de dois estágios

É possível selecionar as seguintes opções para sua análise:

**Salvar novas variáveis.** Permite incluir novas variáveis no arquivo ativo. As opções disponíveis são Preditos e Resíduos.

**Exibir covariância paramétricas.** Permite imprimir a matriz de covariâncias das estimativas de parâmetros.

## Recursos adicionais do comando 2SLS

A linguagem da sintaxe de comando também permite estimar várias equações simultaneamente. Consulte a *Referência de Sintaxe de Comando* para obter informações da sintaxe completa.

## 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 como 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 ... -1/(k-1) -1/(k-1) -1/(k-1))
df(2) ( 0 1 ... -1/(k-2) -1/(k-2) -1/(k-2))
.
.
df(k-2) ( 0 0 ... 1 -1/2 -1/2)
df(k-1) ( 0 0 ... 0 1 -1)
```

em que  $k$  é o número de categorias da variável independente. Por exemplo, uma variável independente com quatro categorias possui uma matriz de contraste de Helmert da seguinte forma:

```
(1/4 1/4 1/4 1/4)
( 1 -1/3 -1/3 -1/3)
( 0 1 -1/2 -1/2)
( 0 0 1 -1)
```

## Diferença

**Diferença ou contrastes de Helmert reversos.** Compara as categorias de uma variável independente com a média das categorias anteriores da variável. O formulário matriz geral é

```
mean ( 1/k 1/k 1/k ... 1/k)
df(1) ( -1 1 0 ... 0)
df(2) ( -1/2 -1/2 1 ... 0)
.
.
df(k-1) (-1/(k-1) -1/(k-1) -1/(k-1) ... 1)
```

em que  $k$  é o número de categorias para a variável independente. Por exemplo, os contrastes de diferença para uma variável independente com quatro categorias são os seguintes:

```
( 1/4 1/4 1/4 1/4)
( -1 1 0 0)
( -1/2 -1/2 1 0)
( -1/3 -1/3 -1/3 1)
```

## Polinomial

**Contrastes 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 contraste polinomiais para executar o ajuste de curva não linear, como regressão curvilinear.

## Repetido

**Compara níveis adjacentes de uma variável independente.** O formulário matriz geral é

```
mean (1/k 1/k 1/k ... 1/k 1/k)
df(1) ( 1 -1 0 ... 0 0)
df(2) ( 0 1 -1 ... 0 0)
      .
      .
df(k-1) ( 0 0 0 ... 1 -1)
```

em que  $k$  é o número de categorias para a variável independente. Por exemplo, os contrastes repetidos para uma variável independente com quatro categorias são os seguintes:

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
( 0 1 -1 0)
( 0 0 1 -1)
```

Esses contrastes são úteis na análise de perfil e sempre que os escores de diferença são necessários.

## Especial

**Um contraste definido pelo usuário.** Permite a entrada de contrastes especiais na forma de matrizes quadradas com tantas linhas e colunas quanto existirem categorias da variável independente fornecida.

Para MANOVA e LOGLINEAR, a primeira linha inserida sempre será a média ou constante, efeito e representa o conjunto de ponderações, indicando como obter a média de outras variáveis independentes, se houver, sobre a variável fornecida. Geralmente, esse contraste é um vetor de uns.

As linhas restantes da matriz 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, exceto as somas de linha média para 0. Produtos de cada par de linhas separadas somam 0 também:

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 4: (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0
```

Os contrastes especiais não precisam ser ortogonais. No entanto, eles não devem ser combinações lineares uns dos outros. Se forem, o procedimento relata a dependência linear e interrompe o processamento. Helmert, diferença e contrastes polinomial são todos contrastes ortogonais.

## Indicador

**Codificação de variável indicadora.** Também conhecido como codificação simulada, esta codificação não está disponível em LOGLINEAR ou MANOVA. O número de novas variáveis codificadas é  $k-1$ . Casos na categoria de referência são codificados como 0 para todas as variáveis  $k-1$ . Um caso na categoria  $i^{\text{th}}$  é codificado como 0 para todas as variáveis indicadoras, exceto o  $i^{\text{th}}$ , que é codificado como 1.

## Regressão de Kernel Ridge

A Regressão de Kernel Ridge é um procedimento de extensão que usa a classe Python **sklearn.kernel\_ridge.KernelRidge** para estimar modelos de regressão de kernel ridge estimados. Os modelos de regressão de kernel ridge são modelos de regressão não paramétricos capazes de modelar relacionamentos lineares e não lineares entre variáveis preditoras e resultados. Os resultados podem ser altamente sensíveis às escolhas de hiperparâmetros de modelo. A Regressão de Kernel Ridge facilita a escolha de valores de hiperparâmetro por meio de validação cruzada de dobra k em grades especificadas de valores usando a classe **sklearn.model\_selection.GridSearchCV**.

**Exemplo:**



## Estatísticas

Additive\_CHI2, CHI2, Cosseno, Laplaciano, Linear, Polinomial, RBF, Sigmoide, Alfa, Gama, Coef0, Grau, validação cruzada, observada versus predita, resíduos versus predito, coeficientes de ponderação dupla, coeficientes de ponderação de espaço do kernel.

## Considerações de dados

### Dados

- É possível especificar qualquer ou todas as oito diferentes funções kernel.
- A função kernel selecionada determina quais hiperparâmetros estão ativos.
- Os hiperparâmetros incluem alfa para regularização de ridge comuns a todos os kernels mais tantos quantos três outros hiperparâmetros para cada função kernel específica.
- Quando vários subcomandos de kernel são especificados, ou mais de um valor para qualquer parâmetro é especificado, uma procura de grade com validação cruzada para avaliar modelos é executada, e o modelo com melhor ajuste com base nos dados mantidos é selecionado.
- A extensão aceita variáveis de divisão do procedimento de Arquivo de Divisão e ponderações usando o procedimento de Casos de Ponderação.
- Quando as ponderações são incluídas, elas são usadas para criar valores ajustados em todas as análises. Devido a limitações no método de escoragem na classe **sklearn.model\_selection.GridSearchCV**, as avaliações de validação cruzada usadas para a seleção de modelo não são ponderadas.

### Suposições

## Obtendo uma Regressão de Kernel Ridge

1. Nos menus, escolha:

**Analisar > Regressão > Kernel Ridge...**

2. Selecione uma variável **Dependente**.
3. Selecione uma ou mais variáveis **Independente**.
4. A configuração padrão **Modelo único** é usada quando apenas um valor para cada parâmetro de função kernel é especificado. Quando a configuração **Modelo único** é selecionada, não é possível especificar funções **Kernel** adicionais e ponderações são totalmente aplicadas em toda a análise, avaliação e pontuação de resultados. Também é possível usar os controles de seta para cima e para baixo para reorganizar as funções kernel.

Opcionalmente, selecione **Seleção de modelo** da lista **Modo**.

Quando **Seleção de modelo** é selecionado da lista **Modo**, é possível incluir múltiplas funções kernel à lista de **Kernels**.

- a. Clique no controle de inclusão (+) para incluir funções adicionais do kernel.
- b. Clique na célula vazia na coluna **Kernel** para selecionar uma função kernel.
- c. Clique duas vezes em qualquer célula de linha de função kernel para especificar valores de parâmetro de função kernel para a coluna correspondente (**Alfa**, **Gamma**, **Coef0**, **Grau**). Para obter informações adicionais, consulte “Parâmetros de kernel” na página 30. Os parâmetros de ajuste da função kernel padrão são listados abaixo.

#### **Additive\_CHI2**

ALPHA=1 GAMMA=1

#### **CHI2**

ALPHA=1 GAMMA=1

#### **Cosseno**

ALPHA=1

**Laplaciano**

ALPHA=1 GAMMA=1/p

**Linear**

A função kernel padrão. ALPHA=1

**Polinomial**

ALPHA=1 GAMMA=1/p COEF0=1 DEGREE=3

**RBF**

ALPHA=1 GAMMA=1/p

**Curva sigmoide**

ALPHA=1 GAMMA=1/p COEF0=1

**Nota:** Quando mais de um valor para qualquer parâmetro de função kernel é especificado, uma procura de grade com validação cruzada para avaliar modelos é executada e o modelo com melhor ajuste baseado nos dados mantidos é selecionado.

5. Opcionalmente, clique em **Opções** para especificar o número de dobras de validação cruzada, opções de exibição, configurações de gráfico e itens a serem salvos. Para obter mais informações, consulte [“Regressão de Kernel Ridge: opções”](#) na página 30.
6. Clique em **OK**.

## Parâmetros de kernel

O diálogo **Parâmetros de Kernel** fornece opções para especificar valores de parâmetro de função de kernel único e para especificar que aquela seleção de modelo seja executada usando uma procura de grade sobre as combinações de kernels e valores de parâmetro de grade especificados.

**Especificar parâmetros únicos**

Ative a configuração para especificar valores para o parâmetro de função do kernel selecionado.

- Insira um valor e clique em **Incluir** para incluir o valor no parâmetro de função do kernel.
- Selecione um valor de parâmetro e clique em **Alterar** para atualizar o valor.
- Selecione um valor de parâmetro e clique em **Remover** para excluir o valor.

**Especificar parâmetros de grade**

Ative a configuração para especificar que a seleção do modelo seja executada usando uma procura de grade nas combinações de kernels e valores de parâmetro de grade especificados.

## Regressão de Kernel Ridge: opções

O diálogo **Gráficos** fornece opções para especificar o número de dobras de validação cruzada, opções de exibição, configurações de gráfico e itens a serem salvos.

**Número de pastas de validação cruzada**

O número de divisões ou dobras na validação cruzada com procura de grade para seleção de modelo. Insira um valor de número inteiro maior que 1. O valor padrão é 5 segundos. A configuração está disponível apenas quando **Seleção de modelo** é escolhida como o **Modo** no diálogo primário **Regressão de Kernel Ridge**.

**Exibir**

Fornece opções para especificar qual saída exibir quando a validação cruzada estiver em vigor.

**Melhor**

A configuração padrão exibe apenas resultados básicos para o melhor modelo escolhido.

**Comparar**

Exibe resultados básicos para todos os modelos avaliados.

**Comparar modelos e pastas**

Exibe resultados completos para cada divisão ou dobra para cada modelo avaliado.

### Gráfico

Fornece opções para especificar gráficos de valores observados ou residuais em relação a valores preditos.

#### Observados vs. preditos

Exibe um gráfico de dispersão de valores observados em relação aos preditos para o modelo especificado ou melhor.

#### Resíduos vs. preditos

Exibe um gráfico de dispersão de valores residuais em relação aos preditos para o modelo especificado ou melhor.

### Salvar

A tabela fornece opções para especificar variáveis para salvar no conjunto de dados ativo.

#### Valores preditos

Salva valores preditos do modelo especificado ou melhor no conjunto de dados ativo. Um nome de variável opcional pode ser incluído.

#### Resíduos

Salva resíduos das predições do modelo especificado ou melhor no conjunto de dados ativo. Um nome de variável opcional pode ser incluído.

#### Coefficientes dual

Salva os coeficientes de ponderação de espaço dupla ou de kernel do modelo especificado no conjunto de dados ativo. Um nome de variável opcional pode ser incluído. A configuração não está disponível quando **Seleção de modelo** é escolhida como o **Modo** no diálogo primário **Regressão de Kernel Ridge**.

## Modelos De Tempo De Falha Acelerada Paramétrica

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

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

1. Nos menus, escolha:

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

2. Selecione uma variável de origem.

#### Hora

##### Sobrevivência

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

##### Início/Término

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

#### Estado

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

##### Falha/Evento

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

##### Censura à direita

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

##### Censura à esquerda

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

### **Censura de intervalo**

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

### **Tratamento de valores não mapeados**

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

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

### **Covariável(is)**

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

### **Fatores fixos**

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

### **Truncamento à esquerda**

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

## **Modelos De Tempo De Falha Acelerada Paramétrica: Critérios**

### **Critérios**

Um painel opcional para especificar os critérios gerais.

### **Intervalo de confiança**

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

### **Valores omissos**

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

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

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

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

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

### **Tratamento de status**

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

#### **Descartar registro conflituado**

Cai os registros conflituados. Esta é a definição padrão.

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

Obtém as informações de tempo de acordo com o status.

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

Altera o status de acordo com as informações do tempo.

## **Modelos De Tempo De Falha Acelerada Paramétrica: Modelo**

### **Modelo**

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

### **Distribuição do tempo de sobrevivência**

Uma opção para especificar a distribuição do tempo de sobrevivência.

#### **Weibull**

Especifica a distribuição Weibull. Esta é a definição padrão.

**Exponencial**

Especifica distribuição exponencial.

**Log-normal**

Especifica log-distribuição normal.

**Logística de log**

Especifica distribuição log-logistic.

**Configurações de covariável**

Especificar variáveis covariadas.

**Configurações do fator**

Especificar variáveis de fator.

**Valor inicial de interceptação**

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

**Valor inicial do parâmetro de escala**

Uma opção para controlar a configuração do parâmetro de escala.

**Erro padrão da regressão OLS correspondente**

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

**Inverte erro padrão da regressão OLS correspondente**

Usa o recíproco do erro padrão.

**Valor fornecido pelo usuário**

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

## Modelos De Tempo De Falha Acelerada Paramétrica: Estimativa

### Estimativa

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

**Método De Direção Alternada ou Multiplicadores (ADMM)****Rápido**

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

**Tradicional**

Aplica o algoritmo ADMM tradicional.

**Aplicar regularização L-1**

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

**Critérios de convergência modelo****Convergência de parâmetro**

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

**Convergência de função objetiva**

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

### **Convergência Hessiana**

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

### **Critérios de convergência residual**

Uma opção para controlar o processo de otimização.

#### **Resíduos primários e duplos**

Aplica critério de convergência residual primal e dual. Esta é a definição padrão.

#### **Apenas residual primário**

Aplica o critério de convergência residual primal.

#### **Apenas residual duplo**

Aplica o critério de convergência residual dual.

### **Método**

Um parâmetro opcional para especificar o método de estimação.

#### **Automático**

Escolhe automaticamente o método baseado no conjunto de dados de amostra. Esse é o padrão. O Campo **Limiar de número de preditores** especifica o limite do número de preditores e deve ser um único inteiro maior que 1. O valor padrão é 1000.

#### **Newton-Raphson**

Aplica o método de Newton-Raphson.

#### **L-BFGS**

Aplica o algoritmo BFGS de memória limitada. O campo **Update** especifica o número das atualizações passadas mantidas pelo algoritmo BFGS de memória limitada, e deve ser um único inteiro maior ou igual a 1. O valor padrão é 5 segundos.

### **Iteração**

#### **Máximo de iterações**

Especifica o número máximo de iterações. Ele deve ser um único inteiro pertencente a  $[1, 100]$ . A configuração padrão é 20.

#### **Máxima divisão de passos pela metade**

Especifica o número máximo de halving de etapa. Ele deve ser um único inteiro pertencente a  $[1, 20]$ . A configuração padrão é 5.

#### **Número máximo de pesquisas de linha**

Especifica o número máximo de pesquisas de linha. Ele deve ser um único inteiro pertencente a  $[1, 100]$ . A configuração padrão é 20.

#### **Convergência absoluta para o processo de iteração**

Especifica a convergência absoluta para o processo de iteração exterior. Deve ser um valor numérico único pertencente a  $(0, 1)$ . A configuração padrão é 0.0001..

#### **Convergência relativa para o processo de iteração**

Especifica a convergência relativa para o processo de iteração externa. Deve ser um valor numérico único pertencente a  $(0, 1)$ . A configuração padrão é 0.01..

## **Modelos De Tempo De Falha Acelerada Paramétrica: Imprimir**

### **Imprimir**

Um painel opcional para controlar as saídas de mesa.

#### **Detalhes de codificação do fator**

Se selecionado, exibe e imprime os detalhes de codificação dos fatores. O processo é ignorado se não houver fatores de efetivação.

### Valores iniciais atribuídos aos parâmetros de regressão

Se selecionado, exibe os valores iniciais utilizados no processo de estimação.

### Histórico de iteração do modelo

Se selecionado, exibe o histórico de iteração de análise de sobrevivência. No campo **Número de etapas**, especifique o número de etapas entre 1 e 99999999. A definição padrão é 1.

### Resultados da seleção contendo

Controla a exibição dos detalhes da seleção do recurso.

#### Ambas as variáveis selecionadas e não selecionadas

Exibir ambas as variáveis selecionadas e não selecionadas na tabela.

#### Apenas variáveis selecionadas

Apenas exibir as variáveis selecionadas.

#### Apenas variáveis não selecionadas

Apenas exibir as variáveis não selecionadas. O campo **Variáveis máximas a exibir** especifica o número máximo das variáveis impressas na tabela. A configuração padrão é 30.

## Modelos De Tempo De Falha Acelerada Paramétrica: Prever

### Prever

Um painel opcional para pontuar e salvar as estatísticas previstas para o conjunto de dados ativos.

### Valores de tempo para escoragem

#### Valores de tempo definidos por variáveis dependentes

Pontue as **Previsões** com base na variável de tempo especificada para o modelo de sobrevivência paramétrica.

#### Intervalos regulares

Pontue as **Previsões** com base em valores de tempo futuros. O campo **Intervalo de tempo** especifica o intervalo de tempo, e deve ser um valor numérico único maior que 0. O campo **Número de períodos de tempo** especifica o número dos períodos de tempo, e deve ser um único inteiro numérico entre 2 e 100.

#### Tempo de duração

Pontue as **Previsões** com base no tempo de duração para definir os valores de tempo futuro. Ele deve ser uma única variável numérica.

### Predições

#### Sobrevivência

Marca e salva as estatísticas de sobrevivência previstas para o conjunto de dados ativos. O nome da variável customizada padrão (ou nome-raiz) é PredSurvival.

#### Risco

Marca e salva os perigos previstos para o conjunto de dados ativos. O nome da variável customizada padrão (ou nome-raiz) é PredHazard.

#### Risco acumulativo

Pontue e salva os perigos acumulativos previstos para o conjunto de dados ativos. O nome da variável customizada padrão (ou nome-raiz) é PredCumHazard.

#### Sobrevivência condicional

Marca e salva as estatísticas de sobrevivência condicional previstas para o conjunto de dados ativos. O nome da variável customizada padrão (ou nome-raiz) é PredConditionalSurvival. O processo será ignorado se PASTTIME não for especificado. Um valor **Tempo de sobrevivência passado** é necessário e especifica os valores de tempo passado para escoragem. Ele deve ser uma única variável numérica.

## Modelos De Tempo De Falha Acelerada Paramétrica: Plot

### Gráfico

#### Gráfico de resíduos de Cox-Snell

Selecione **Exibir o gráfico** para criar um gráfico residual Cox -Snell. No campo **Número de pontos de corte de binning**, especifique um número de 1 10000. A configuração padrão é 100.

#### Gráficos de função

Uma opção para controlar as tramas de função.

##### Tipo

###### Sobrevivência

Cria a trama para funções de sobrevivência.

###### Risco

Cria a trama para as funções de risco.

###### Densidade

Cria uma trama para as funções de densidade.

#### Número de pontos a serem exibidos

Especifica o número de pontos de função entre 1 e 200. A configuração padrão é 100.

#### Valores covariados para gráfico

Um opcional para especificar os valores fornecidos pelo usuário e atribuí-los aos preditores. Por padrão, as parcelas designadas serão criadas no **Mean** de cada covariado em efeito e na frequência de categoria de cada fator em vigor. Se especificado, as parcelas designadas serão criadas com base na configuração do padrão. Na presença de quaisquer variáveis duplicadas, o especificado primeiro seria reconhecido e o restante seria ignorado. Uma variável válida deve ser contida em um efeito modelo. Para um covariado, o valor fornecido pelo usuário deve ser numérico. Omissão de uma variável em efeito indica que a frequência da categoria e o **Mean** seriam usados por padrão para o fator e o covariado, respectivamente. Se um valor inválido for atribuído a uma variável, o padrão solicitado não será plotado.

#### Valores de fator para gráfico

Um opcional para especificar os valores fornecidos pelo usuário e atribuí-los aos preditores. Na presença de quaisquer variáveis duplicadas, o especificado primeiro seria reconhecido e o restante seria ignorado. Uma variável válida deve ser contida em um efeito modelo. Omissão de uma variável em efeito indica que a frequência da categoria e a média seriam utilizadas por inadimplência para o fator e o covariado, respectivamente. Se um valor inválido for atribuído a uma variável, o padrão solicitado não será plotado.

#### Linhas separadas para

Uma opção para especificar uma variável categórica pela qual as parcelas da linha serão desenhadas.

#### Número máximo de linhas em um gráfico

Especifica o número máximo das linhas em um gráfico se **Separar linhas para** for especificado. A configuração padrão é 10.

## Modelos De Tempo De Falha Acelerada Paramétrica: Exportação

### Exportar

Selecione **Exportar informações do modelo para arquivo XML** para gravar as informações do modelo e do parâmetro em um arquivo PMML para pontuação. Você deve especificar o diretório e o nome do arquivo do arquivo PMML a ser salvo.



# Sobrevivência AFT Definir Eventos para Variáveis de Status

Ocorrências do valor ou valores selecionados para a variável de status indicam que o evento terminal ocorreu para esses casos. Todos os outros casos são considerados como censurados. Insira um valor único ou um intervalo de valores que identifica o evento de interesse.

## Modelos de Tempo de Falha Acelerada Paramétrica: Selecionar Categoria

A configuração Selecionar Categoria fornece opções para escolher um valor que indica a categoria a ser modelada como uma linha de base para comparação.

Selecionando a categoria

Clique em 'Última Categoria' para abrir a caixa de diálogo 'Selecionar categoria'.

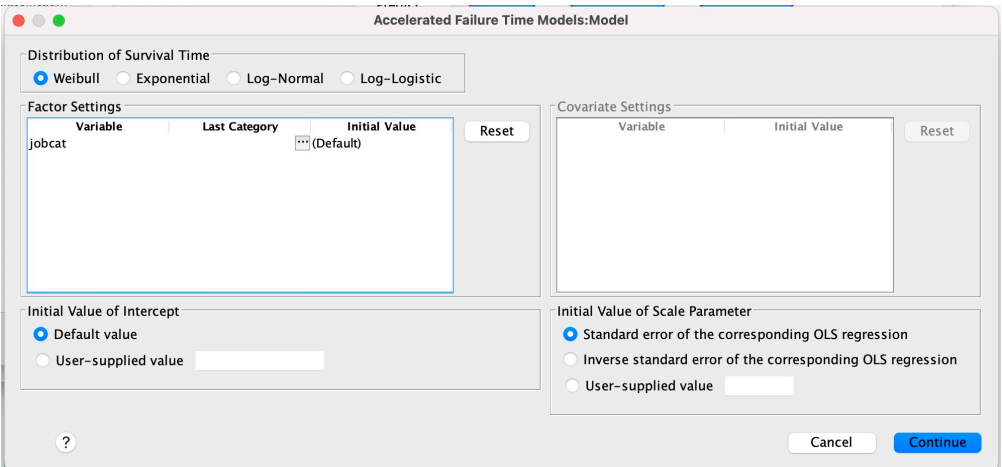


Figura 1. Modelos de Tempo de Vida Acelerado-Caixa de Diálogo-Categoria

Para designar uma categoria como a linha de base, selecione um valor na caixa de diálogo 'Selecionar categoria'.

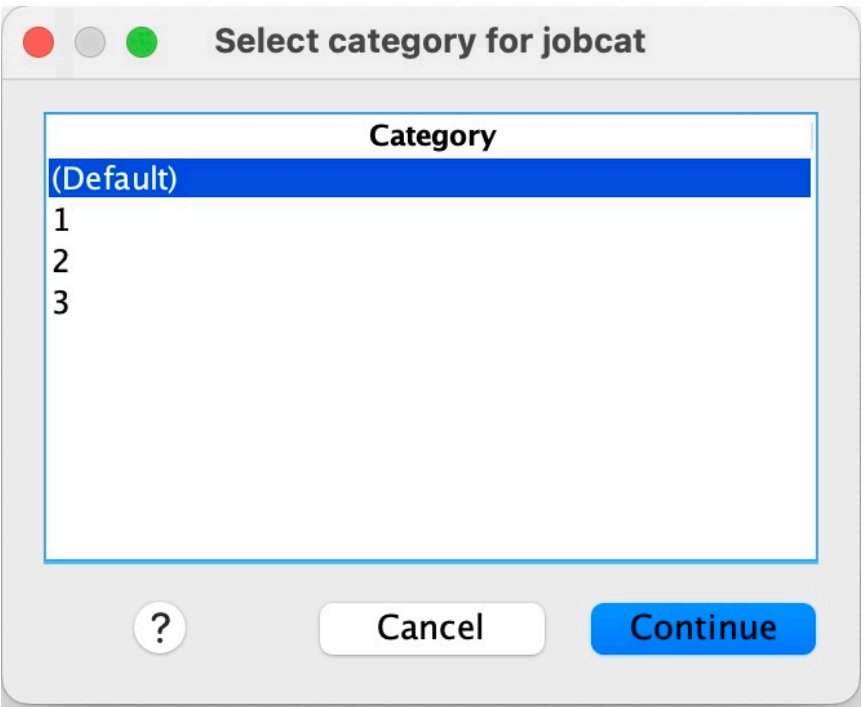


Figura 2. Modelos de Tempo de Vida Acelerado-Caixa de Diálogo-Selecionar categoria

Clique em Continuar.

## Modelos de fragilidade compartilhados paramétricos

---

A Parametric Shared Frailty Models Survival analysis inicia o procedimento paramétrico de modelos de sobrevivência com entrada de dados de tempo de vida recorrente. Os modelos paramétricos de sobrevivência assumem que o tempo de sobrevivência segue uma distribuição conhecida, e esta análise incorpora um termo de fragilidade em um modelo paramétrico de sobrevivência. É tratado como um componente aleatório para dar conta de um efeito não observado devido à variabilidade de nível individual ou de grupo.

### Obtendo uma análise paramétrica de Modelos de Fragilidade Compartilhados

1. No menu, escolha:

**Analisar > Sobrevivência > Modelos Paramétricos Compartilhados de Fragilidade**

2. Selecione uma variável de origem.

#### Hora

##### Sobrevivência

O tempo de sobrevivência é representado por uma variável para indicar o horário de encerramento. O horário de início seria configurado como 0

##### Início / Fim

Variáveis numéricas que denotam **Horário de Início** e **Horário de Encerramento**.

#### Assunto

Necessário para executar o procedimento.. Especifica uma única variável para o ID do sujeito..

#### Intervalo

Especifica uma variável única e numérica para o número de intervalo usado para identificar os diferentes registros recorrentes que compartilham o mesmo ID de sujeito.

#### Estado

Sequência opcional única ou variável numérica que determina uma das seguintes configurações de status:

##### Falha/Evento

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

##### Censura à direita

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

##### Tratamento de valores não mapeados

Controla para qual categoria mapear os registros não mapeados.. Para excluir os registros que falharam ao serem mapeados, selecione **Excluir deles da análise**.

Clique no botão **Definir evento** para definir um evento para a variável de status

#### Covariável(is)

Uma ou mais variáveis numéricas opcionais a serem tratadas como covariáveis Observe que uma variável não pode ser especificada por **Covariáveis** e **Fator (s) Fixo (s)**.

#### Fatores fixos

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

## Modelos de Fragilidade Compartilhados Paramétricos: Critérios

### Critérios

Um painel opcional para especificar os critérios gerais

#### Intervalo de confiança

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

#### Nível de Significância

Uma opção para especificar o nível de significância do teste de razão de verossimilhança para o componente de fragilidade. Ele deve ser um valor numérico único entre 0 e 1. A configuração padrão é 0.05..

#### Valores omissos

Uma opção para controlar como os valores omissos de usuário são tratados:

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

Trata os valores omissos de usuário como valores válidos.. Esse é o padrão.

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

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

#### Tratamento de intervalo

Uma opção para controlar como lidar com registros cujo intervalo entra em conflito com o horário de início e de término. Ele entrará em vigor se houver duas variáveis de tempo com uma variável Intervalo especificada no diálogo principal.

##### Descartar os registros com conflitos

Descarta os registros seriais do sujeito inteiro se o valor do intervalo estiver em conflito com o horário de início e de término. Esta é a definição padrão.

##### Descobrir valores de intervalo com base no horário de início e término

Descobre o valor do intervalo do horário de início e de encerramento.

## Modelos de fragilidade compartilhados paramétricos: Modelo

### Modelo

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

#### Distribuição do tempo de sobrevivência

Uma opção para especificar a distribuição do tempo de sobrevivência.

##### Weibull

Especifica a distribuição de Weibull. Esta é a definição padrão.

##### Exponencial

Especifica a distribuição exponencial.

##### Log-normal

Especifica a distribuição de Log-normal

##### Logística de log

Especifica a distribuição logística de log

#### Configurações de covariável

Especifique variáveis covariáveis.

#### Configurações do fator

Especifique as variáveis de fator

#### Valor inicial de interceptação

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

**Valor inicial do parâmetro de escala**

Uma opção para controlar a configuração do parâmetro de escala.

**Erro padrão da regressão OLS correspondente**

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

**Inverte erro padrão da regressão OLS correspondente**

Usa a recíproca do erro padrão

**Valor fornecido pelo usuário**

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

**Componente de fragilidade**

Um parâmetro opcional para especificar a **Distribuição** do componente de fragilidade

**Gama**

Especifica a distribuição gama. Esta é a definição padrão.

**Inverso-Gaussiano**

Especifica a distribuição inversa-gaussiana

**Valor inicial da variância**

Especifica o valor inicial da variação do componente de fragilidade. Deve ser um valor numérico único maior que 0. O valor padrão é 1.0 para distribuição Gama e 0.1 para distribuição Gaussiana inversa.

## Modelos Paramétricos de Fragilidade Compartilhados: Estimativa

**Estimativa**

Um painel opcional para especificar as configurações para controlar a estimação dos modelos de fragilidade compartilhados e o processo de seleção de recurso opcional

**Método ou Multiplicadores de Direção Alternados (ADMM)****Rápido**

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

**Tradicional**

Aplica o algoritmo tradicional do ADMM

**Aplicar regularização L-1**

Conduz o processo para controlar a seleção de variável O campo **Parâmetro de Penalidade** especifica o parâmetro de penalidade que controla o processo de regularização.. Deve ser um valor único maior que 0. A configuração padrão é 0.001..

**Critérios de convergência modelo****Convergência de parâmetro**

Especifica os critérios de convergência para o parâmetro Ele deve ser um valor numérico único pertencente a  $[0, 1)$ . A configuração padrão é 0.000001. Para **Tipo**, é possível selecionar **ABSOLUTE** para aplicar a convergência absoluta à otimização interna ou **RELATIVE** para aplicar a convergência relativa à otimização interna. O **Valor** opcional especifica um limite numérico para o tipo de convergência.

**Convergência de função objetiva**

Especifica os critérios de convergência para a função objetivo. Ele deve ser um valor numérico único que pertence a  $[0, 1)$ . A configuração padrão é 0, que não aplica os critérios de convergência. Para **Tipo**, é possível selecionar **ABSOLUTE** para aplicar a convergência absoluta à otimização interna ou **RELATIVE** para aplicar a convergência relativa à otimização interna. O **Valor** opcional especifica um limite numérico para o tipo de convergência.

### **Convergência Hessiana**

Especifica os critérios de convergência para a matriz Hessiana. Ele deve ser um valor numérico único que pertence a  $[0, 1)$ . A configuração padrão é 0, que não aplica os critérios de convergência. Para **Tipo**, é possível selecionar **ABSOLUTE** para aplicar a convergência absoluta à otimização interna ou **RELATIVE** para aplicar a convergência relativa à otimização interna. O **Valor** opcional especifica um limite numérico para o tipo de convergência.

### **Critérios de convergência residual**

Uma opção para controlar o processo de otimização.

#### **Resíduos primários e duplos**

Aplica o critério de convergência residual primário e dual. Essa configuração é por padrão.

#### **Apenas residual primário**

Aplica o critério de convergência residual primordial.

#### **Apenas residual duplo**

Aplica o critério de convergência residual dual.

### **Método**

Um parâmetro opcional para especificar o método de estimação

#### **Automático**

Escolhe automaticamente o método baseado no conjunto de dados de amostra. Esse método é selecionado por padrão.. O campo **Número limite de preditores** especifica o limite do número de preditores e deve ser um número inteiro único maior que 1. O valor padrão é 1000.

#### **Newton-Raphson**

Aplica o método de Newton-Raphson.

#### **L-BFGS**

Aplica o algoritmo de BFGS de memória limitada O campo **Atualizar** especifica o número de atualizações passadas que são mantidas pelo algoritmo BFGS de memória limitada e deve ser um número inteiro maior ou igual a 1. O valor padrão é 5 segundos.

### **Iteração**

#### **Máximo de iterações**

Especifica o número máximo de iterações. Ele deve ser um número inteiro único que pertence a  $[1, 300]$ . A configuração padrão é 20.

#### **Máxima divisão de passos pela metade**

Especifica o número máximo de redução da etapa pela metade Ele deve ser um número inteiro único que pertence a  $[1, 200]$ . A configuração padrão é 5.

#### **Número máximo de procuras de linha**

Especifica o número máximo de procuras de linhas. Ele deve ser um número inteiro único que pertence a  $[1, 300]$ . A configuração padrão é 20.

#### **Convergência absoluta para o processo de iteração**

Especifica a convergência absoluta para o processo de iteração externa Ele deve ser um único valor numérico que pertence a  $(0, 1)$ . A configuração padrão é 0.0001..

#### **Convergência relativa para processo de iteração**

Especifica a convergência relativa para o processo de iteração externa Ele deve ser um único valor numérico que pertence a  $(0, 1)$ . A configuração padrão é 0.01..

## **Modelos de fragilidade compartilhados paramétricos: Impressão**

### **Imprimir**

Um painel opcional que controla as saídas da tabela

#### **Detalhes de codificação do fator**

Se selecionado, exibe e imprime os detalhes de codificação dos fatores O processo será ignorado se não houver fatores em vigor..

### Valores iniciais que são designados aos parâmetros de regressão

Se selecionado, exibe os valores iniciais usados no processo de estimação.

### Histórico de iteração do modelo

Se selecionado, exibe o histórico de iteração da análise de sobrevivência. No campo **Número de etapas**, especifique o número de etapas entre 1 e 99999999. A definição padrão é 1.

## Modelos de fragilidade compartilhados paramétricos: prever

### Prever

Um painel opcional para pontuar e salvar as estatísticas preditas para o conjunto de dados ativo

#### Valores de tempo para escoragem

##### Valores de tempo definidos por variáveis dependentes

Pontua as **Predições** com base na variável de tempo especificada para o modelo de sobrevivência paramétrica..

##### Intervalos regulares

Pontua as **Predições** com base em valores de tempo futuros. O campo **Intervalo de Tempo** especifica o intervalo de tempo e deve ser um valor numérico único maior que 0. O campo **Número de períodos de tempo** especifica o número dos períodos de tempo e deve ser um número inteiro numérico único entre 2 e 100.

##### Tempo de duração

Pontua as **Predições** com base na duração de tempo para definir os valores de tempo futuros. Ela deve ser uma única variável numérica.

### Predições

#### Sobrevivência

Pontua e salva as estatísticas de sobrevivência previstas no conjunto de dados ativo. O nome da variável customizada padrão (ou nome-raiz) é `PredSurvival`.

#### Risco

Pontua e salva os perigos previstos no conjunto de dados ativo. O nome da variável customizada padrão (ou nome-raiz) é `PredHazard`.

#### Risco acumulativo

Pontua e salva os perigos acumulativos previstos no conjunto de dados ativo. O nome da variável customizada padrão (ou nome-raiz) é `PredCumHazard`.

#### Sobrevivência condicional

Escores e salva as estatísticas de sobrevivência condicional predita para o conjunto de dados ativo. O nome da variável customizada padrão (ou nome-raiz) é `PredConditionalSurvival`. O processo será ignorado se `PASTTIME` não for especificado. Um valor **Tempo de sobrevivência passado** é necessário e especifica os valores de tempo passado para escoragem. Ela deve ser uma única variável numérica.

#### Sobrevida incondicional

Escore e salva as estatísticas de sobrevivência incondicional previstas no conjunto de dados ativo. A palavra-chave é suprimida por padrão.. Se especificado, ele poderia ser seguido por um nome de variável opcional fornecido pelo usuário (ou nome raiz) especificado entre parênteses. O nome padrão é `PredUnCondSurvival`..

#### Risco incondicional

Escores e salva as estatísticas de risco incondicional predito para o conjunto de dados ativo. A palavra-chave é suprimida por padrão.. Se especificado, ele poderia ser seguido por um nome de variável opcional fornecido pelo usuário (ou nome raiz) especificado entre parênteses. O nome padrão é `PredUncondHazard`..

#### Risco de cum incondicional

Escores e salva as estatísticas de risco acumulativo incondicional predito no conjunto de dados ativo. A palavra-chave é suprimida por padrão.. Se especificado, ele poderá ser seguido por um

nome de variável opcional fornecido pelo usuário (ou nome raiz especificado entre parênteses. O nome padrão é PredUncondCumHazard..

## Modelos de Fragilidade Compartilhados Paramétricos: Plot

### Gráfico

#### Gráficos de função

Uma opção para controlar os gráficos de função

##### Tipo

###### Sobrevivência

Cria o gráfico para as funções de sobrevivência incondicional.

###### Risco

Cria o gráfico para as funções de risco incondicional.

###### Densidade

Cria um gráfico para as funções de densidade..

#### Número de pontos a exibir

Especifica o número de pontos de função entre 1 e 200 A configuração padrão é 100.

#### Valores covariados para gráfico

Uma opção para especificar valores fornecidos pelo usuário e designá-los aos preditores. Por padrão, os gráficos designados serão criados na Média de cada covariável em vigor.. Se especificado, os gráficos designados serão criados com base na configuração do padrão.. Na presença de quaisquer variáveis duplicadas, o especificado primeiro seria reconhecido e o restante seria ignorado.. Uma variável válida deve estar contida em um efeito de modelo Para uma covariável, o valor fornecido pelo usuário deve ser numérico. A omissão de uma variável em efeito indica que a Média seria usada por padrão para a covariável.. Se um valor inválido for designado a uma variável, o padrão solicitado não será plotado.

#### Valores de fator para gráfico

Uma opção para especificar valores fornecidos pelo usuário e designá-los aos preditores. Por padrão, os gráficos designados serão criados na frequência de categoria de cada fator em vigor Se especificado, os gráficos designados serão criados com base na configuração do padrão.. Na presença de quaisquer variáveis duplicadas, o especificado primeiro seria reconhecido e o restante seria ignorado.. Uma variável válida deve estar contida em um efeito de modelo A omissão de uma variável em efeito indica que a frequência da categoria seria usada por padrão para o fator.. Se um valor inválido for designado a uma variável, o padrão solicitado não será plotado.

#### Linhas separadas para

Uma opção para especificar uma variável categórica pela qual os gráficos de linha serão desenhados.

#### Número máximo de linhas em um gráfico

Especifica o número máximo das linhas em um gráfico se **Separar linhas para** for especificado A configuração padrão é 10.

## Modelos paramétricos compartilhados de fragilidade: Exportação

### Exportar

Selecione **Exportar informações de modelo para arquivo XML** para gravar as informações de modelo e parâmetro em um arquivo PMML para escoragem. Você deve especificar o diretório e o nome do arquivo PMML a ser salvo.

## Modelos de fragilidade compartilhados paramétricos: definir eventos

Uma opção para definir status. Se a variável de status for omitida, falha ou evento se tornará o status padrão para todos os casos.

1. A partir do menu escolher,

**Analisar > Sobrevivência > Modelos Paramétricos Compartilhados de fragilidade ...**

- 2.

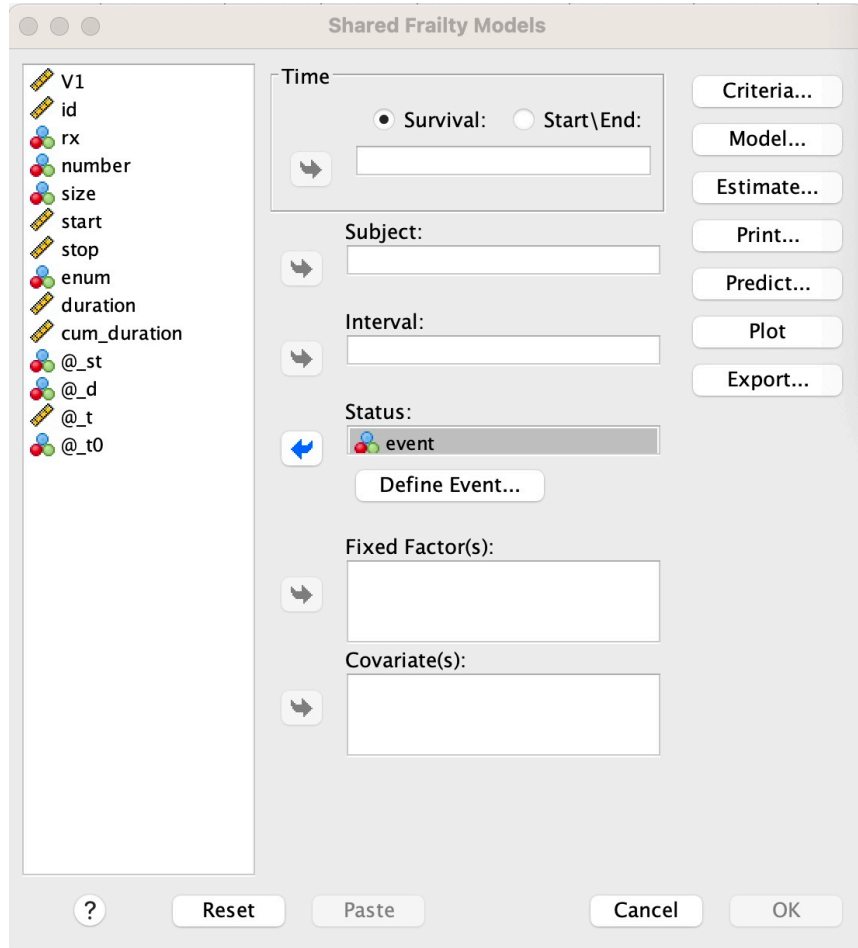


Figura 3. Modelos de fragilidade compartilhados-caixa de diálogo-Status

- 3.

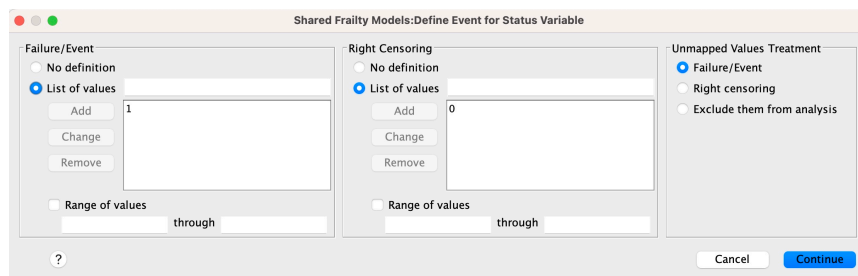


Figura 4. Modelos de fragilidade compartilhados-caixa de diálogo-Status-Definir evento

## Modelos de fragilidade compartilhados paramétricos-Exemplos

### Exemplo 1

SURVREG RECORRENTE y COM x1 BY x2

/MODEL SUBJECT = id FRAILTY=GAMA DISTRIBUTION=WEIBULL.



Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de  $y$  em uma covariável  $x_1$  e no fator  $x_2$ .

O tempo de sobrevivência é representado por uma única variável  $y$ .

Os sujeitos são identificados pela variável  $id$ .

O tempo de sobrevivência é assumido para seguir uma distribuição Weibull.

Assume-se que a variância da fragilidade segue uma distribuição Gama.

Todos os registros válidos são usados na análise de sobrevida.

### **Exemplo 2**

SURVREG RECORRENTE  $y$  COM  $x_1$  BY  $x_2$

/MODEL SUBJECT =  $id$  FRAILTY=INV\_GAUSSIAN DISTRIBUTION=LOG\_NORMAL INTERVAL= $z$ .

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de  $y$  em uma covariável  $x_1$  e no fator  $x_2$ .

O tempo de sobrevivência é representado por duas variáveis  $y_1$  e  $y_2$  denotando o horário de início e de encerramento.

Os sujeitos são identificados pela variável  $id$ .

Presume-se que o tempo de sobrevivência siga uma distribuição Log-normal

Assume-se que a variância da fragilidade segue uma distribuição Gaussiana inversa.

Intervalos de tempo são definidos pela variável  $z$ . Para cada sujeito, o procedimento usa apenas os registros não conflitantes e exclui da análise todos os registros após o primeiro status de falha

### **Exemplo 3**

SURVREG RECORRENTE  $y_1$   $y_2$  WITH  $x_1$  BY  $x_2(1)$

/MODEL SUBJECT =  $id$  FRAILTY=INV\_GAUSSIAN DISTRIBUTION=LOG\_LOGISTIC

/STATUS VARIABLE=evento FAILURE=1 RIGHT=0.

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de  $y$  em uma covariável  $x_1$  e no fator  $x_2$ . Sobrevivência

é representado por duas variáveis  $y_1$  e  $y_2$  denotando o horário de início e de encerramento, respectivamente. Para o fator  $x_2$ , a categoria "1" é designada como uma linha de base a ser modelada.

Os sujeitos são identificados pela variável  $id$ .

O tempo de sobrevivência é assumido para seguir uma distribuição Log-logística.

Assume-se que a variância da fragilidade segue uma distribuição Gaussiana inversa.

O evento de variável é especificado para definir o status com 1 e 0 denotando falha e censura direita, respectivamente.

### **Exemplo 4**

SURVREG RECORRENTE  $y$  COM  $x_1$  BY  $x_2$

/ASSUNTO DO MODELO =  $id$

/STATUS VARIABLE=evento FAILURE=1 RIGHT=0

/PREVER UNCONDSURVIVAL UNCONDHAZARD UNCONDCUMHAZARD

/FUNCTIONPLOT SURVIVAL HAZARD DENSITY PLOTBY ( $x_2$ ).

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de y em uma covariável x1 e no fator x2. O tempo de sobrevivência é representado por uma única variável y.

Os sujeitos são identificados pela variável id.

Sobrevivência incondicional ou baseada em população, risco e risco acumulativo são pontuados e salvos no conjunto de dados ativo.

As curvas de sobrevivência e de risco incondicionais ou baseadas em população são plotadas separadas pelas categorias em x2.

### **Exemplo 5**

SURVREG RECORRENTE y COM x1 BY x2

/MODEL SUBJECT = id FRAILTY=GAMMA DISTRIBUTION=WEIBULL

/STATUS VARIABLE=evento FAILURE=1 RIGHT=0

/ESTIMATION HCONVERGE=1e-12(RELATIVE ) PCONVERGE=0 FCONVERGE=0SELECTFEATURES=TRUE PENALTY=0.01.

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de y em uma covariável x1 e no fator x2. O tempo de sobrevivência é representado por uma única variável y.

Os sujeitos são identificados pela variável id.

O tempo de sobrevivência é assumido para seguir uma distribuição Weibull.

Assume-se que a variância da fragilidade segue uma distribuição Gama.

Os critérios de convergência são baseados na matriz Hessiana. Ele usa 1e-12 como convergência relativa.

O modelo inclui um termo de penalidade para controlar o processo de regularização O parâmetro de multa é configurado como 0.01.

### **Exemplo 6**

SURVREG RECORRENTE y COM x1 BY x2

/ASSUNTO DO MODELO = id

/STATUS VARIABLE=infetar FAILURE=1 RIGHT=0

/ESTIMATION MAXLINESEARCH=100 MAXITER=50 MAXSTEPHALVING=20

Um modelo paramétrico de sobrevivência de fragilidade compartilhada é ajustado de y em uma covariável x1 e no fator x2. O tempo de sobrevivência é representado por uma única variável y.

Os sujeitos são identificados pela variável id.

O procedimento especifica o número máximo da procura de linha para ser 100, o número máximo de iterações para ser 50 e o número máximo de redução da etapa para ser 20

## **Modelos paramétricos compartilhados de fragilidade-um estudo de caso para dados recuperados**

Modelos paramétricos compartilhados de fragilidade-um estudo de caso para dados recuperados

Nome do caso de uso-Efeito colateral do tratamento.

Atores-Investigador e profissional de saúde pública.

Pré-condições-Um conjunto de dados limpo disponível com base no tempo de sobrevivência, status do efeito secundário e preditores a serem ajustados.

Descrição-Patrick, investigador de saúde pública, está investigando uma amostra de dados que inclui 20 participantes. Estes participantes são recrutadas num estudo sobre um efeito secundário ligeiro que é potencialmente causado por um novo tratamento. O designer de tratamento afirma que não haveria diferenças entre homens e mulheres, em relação ao efeito colateral. Patrick gostaria de avaliar tal hipótese. As variáveis incluídas na amostra de dados são listadas da seguinte forma:

- patID: Número de identificação para identificar um único participante.
- endTime: Tempo de sobrevivência (em dias) do efeito colateral após um tratamento, medido desde o início do tratamento até o relato de um efeito colateral ou a censura em 60 dias.
- sideEffect: Status do efeito colateral, status = 0 se censurado e status = 1 se o efeito colateral leve for relatado.
- idade: idade do participante no período da pesquisa.
- fêmea: fêmea = 0 se macho e fêmea = 1 se fêmea.

Diversos tratamentos podem ser aplicados, o que resulta em diversos registros de tempos de recorrência medidos para um determinado participante. O horário de início é sempre 0 para cada registro, que é omitido na amostra de dados. Patrick está interessado em visualizar as funções de sobrevivência e risco para desenhar uma comparação entre um macho e uma fêmea, controlando sua idade e fragilidade. Ele está ciente de que os tratamentos que são administrados ao mesmo participante são mais correlacionados. Ao assumir que o tempo de sobrevivência segue uma distribuição Weibull, Patrick decide construir um modelo paramétrico de sobrevivência de fragilidade compartilhada no SPSS Statistics para considerar a dependência de tratamento para o mesmo participante.

Sintaxe-

```
DATA LIST FREE
/patID(F5.0) endTime(F5.0) sideEffect(F2.0) age(F5.2) female(F2.0) .
BEGIN DATA .
1 45 0 38.00 0
2 26 1 20.00 1
3 58 0 53.00 0
4 31 1 37.00 1
4 24 0 37.00 1
4 50 0 37.00 1
5 20 1 51.00 0
5 38 1 51.00 0
6 30 0 35.00 1
7 22 1 58.00 1
8 53 1 29.00 1
8 49 1 29.00 1
9 25 0 45.00 0
9 25 0 45.00 0
10 27 0 33.00 1
11 34 1 21.00 1
11 40 0 21.00 1
11 49 0 21.00 1
12 42 1 26.00 0
13 25 0 40.00 0
14 21 1 52.00 0
14 32 1 52.00 0
15 56 0 28.00 1
15 34 0 28.00 1
16 30 0 41.00 0
16 29 0 41.00 0
17 25 1 27.00 0
18 26 1 54.00 1
18 36 1 54.00 1
19 27 0 39.00 0
20 58 1 22.00 1
20 54 0 22.00 1
20 43 1 22.00 1
END DATA.
SURVREG RECURRENT endTime WITH age BY female
/MODEL SUBJECT=patID FRAILTY=GAMMA DISTRIBUTION=WEIBULL
/ESTIMATION HCONVERGE=1e-12 PCONVERGE=0 FCONVERGE=0
/STATUS VARIABLE=sideEffect FAILURE=1 RIGHT=0
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY(female) .
```

**Sinopse:**

A sintaxe especificada por Patrick designa endTime como uma única variável de tempo dependente. O procedimento assume automaticamente que o horário de início é 0 para cada registro.. As variáveis idade e sexo feminino são modeladas como covariável e fator, respectivamente. Supõe-se que os tempos de sobrevivência de recorrência seguem uma distribuição Weibull. Presume-se que o termo fragilidade não observado segue uma distribuição Gamma e seu componente de variância é modelado. Em relação às saídas, a tabela Resumo do Modelo fornece as informações de procedimento e modelo. A tabela Resumo de Processamento de Caso fornece uma listagem abrangente do status de falha / censura e também os casos que são excluídos da análise

Na amostra dos dados de Patrick, todos os registros são válidos e incluídos na análise. Ao comparar o log da verossimilhança com aquele do modelo correspondente sem o componente de fragilidade, o modelo de fragilidade compartilhada falha ao atingir um nível significativo (valor  $p = 0.168$ ). Patrick está se perguntando se é necessário incluir um termo de fragilidade compartilhada no modelo. O fator de aceleração estimado de um participante masculino é 1.017, que é obtido calculando o expoente do coeficiente de regressão estimado 0.017 de [feminino = 0.0]. Seu intervalo de confiança de 95% associado é (.688, 1.504) Esses resultados sugerem que um indivíduo do sexo masculino tem quase o mesmo fator de aceleração que um indivíduo do sexo feminino com a mesma idade e fragilidade. No nível de população, Patrick plota as curvas de sobrevivência incondicional e de risco separadamente para homens e mulheres que são avaliados na média de idade da amostra (37.45 anos).

Patrick confirma que, para qualquer valor fixo de tempo de sobrevivência, um homem e uma mulher, em média, devem ter a mesma probabilidade de sobrevivência. Curiosamente, apesar de uma forma unimodal que é mostrada no gráfico de risco incondicional, Patrick descobre que dentro de um período de 60 dias o risco da população está realmente aumentando. Esse comportamento pode implicar a existência do efeito fragilidade. Para investigar melhor o efeito colateral causado pelos tratamentos, Patrick pode continuar com um modelo sem o componente de fragilidade e comparar o comportamento de homens e mulheres. Além disso, ele pode considerar o acompanhamento com os participantes por um período superior a 60 dias para coletar mais dados.

## Avisos

---

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

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

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

*Gerência de Relações Comerciais e Industriais da IBM Brasil*  
*IBM Brasil Ltda*  
*Botafogo*  
*Rio de Janeiro, RJ*  
*CEP 22290-240*

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

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

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

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

Todas as referências nestas informações para websites não IBM são fornecidas apenas para conveniência e não de forma a servir como um endosso a esses websites. Os materiais contidos nesses websites não fazem parte dos materiais desse produto IBM e a utilização desses websites é de inteira responsabilidade do Cliente.

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

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

*Gerência de Relações Comerciais e Industriais da IBM Brasil*  
*IBM Brasil Ltda*

*Botafogo  
Rio de Janeiro, RJ  
CEP 22290-240*

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

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

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

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

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

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

#### LICENÇA DE COPYRIGHT:

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

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

© Copyright IBM Corp. 2021. Partes deste código são derivadas de Programas de Amostra da IBM Corp. Programas de amostra.

© Copyright IBM Corp. 1989 - 2021. Todos os direitos reservados.

## Marcas comerciais

---

IBM, o logotipo IBM e [ibm.com](http://ibm.com) são marcas comerciais ou marcas registradas da International Business Machines Corp., registradas em várias jurisdições no mundo todo. Outros nomes de produto e de serviço podem ser marcas registradas da IBM ou de outras empresas. A lista atual de marcas comerciais da IBM está disponível na web em "Copyright and trademark information" em [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

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

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

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

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

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

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





# Índice remissivo

## A

Análise de probitos  
  critérios [11](#)  
  definir intervalo [11](#)  
  estatísticas [11](#)  
  intervalos de confiança fiduciários [11](#)  
  iterações [11](#)  
  potência média relativa [11](#)  
  recursos adicionais do comando [11](#)  
  taxa de resposta natural [11](#)  
  teste de paralelismo [11](#)  
análise de regressão logística [2](#)  
análise de sobrevivência  
  em regressão de Kernel Ridge [28](#)

## C

categoria de referência  
  na Regressão logística multinomial [7](#)  
células com zero observações  
  na Regressão logística multinomial [8](#)  
contrastes  
  em Regressão logística [3](#)  
covariáveis  
  em Regressão logística [3](#)  
covariáveis categóricas [3](#)  
covariáveis de sequência de caracteres  
  em Regressão logística [3](#)  
critério de convergência  
  na Regressão logística multinomial [8](#)

## D

dados  
  na Regressão logística multinomial [6](#)  
delta  
  como correção para as células com zero observações [8](#)  
DfBeta  
  em Regressão logística [4](#)

## E

eliminação backward  
  em Regressão logística [3](#)  
estatísticas de qualidade de ajuste de Hosmer-Lemeshow  
  em Regressão logística [5](#)  
Estimação Ponderada  
  estatísticas [22](#)  
  exemplo [22](#)  
  exibir ANOVA e estimativas [23](#)  
  histórico de iteração [23](#)  
  log da verossimilhança [22](#)  
  recursos adicionais do comando [23](#)  
  salvar melhores ponderações como nova variável [23](#)  
estimativas de parâmetro

estimativas de parâmetro (*continuação*)  
  na Regressão logística multinomial [7](#)  
etapa pela metade  
  na Regressão logística multinomial [8](#)

## F

função deviance  
  para estimar o valor de escala de dispersão [8](#)

## H

histórico de iteração  
  na Regressão logística multinomial [8](#)

## I

ID de Cook  
  em Regressão logística [4](#)  
intervalos de confiança  
  na Regressão logística multinomial [7](#)  
intervalos de confiança fiduciários  
  em Análise de probit [11](#)  
iterações  
  em Análise de probit [11](#)  
  em Regressão logística [5](#)  
  na Regressão logística multinomial [8](#)

## L

lei dos retornos decrescentes de Metcherlich  
  na Regressão não linear [19](#)  
log da verossimilhança  
  em Estimação ponderada [22](#)  
  na Regressão logística multinomial [7](#)

## M

matriz de correlações  
  na Regressão logística multinomial [7](#)  
matriz de covariâncias  
  na Regressão logística multinomial [7](#)  
modelo de densidade  
  na Regressão não linear [19](#)  
modelo de densidade de lucro  
  na Regressão não linear [19](#)  
modelo de Gauss  
  na Regressão não linear [19](#)  
modelo de Gompertz  
  na Regressão não linear [19](#)  
modelo de Johnson-Schumacher  
  na Regressão não linear [19](#)  
modelo de Michaelis Menten  
  na Regressão não linear [19](#)  
modelo de Morgan-Mercer-Florin

- modelo de Morgan-Mercer-Florin (*continuação*)  
na Regressão não linear [19](#)
- modelo de Peal-Reed  
na Regressão não linear [19](#)
- modelo de razão de cúbicos  
na Regressão não linear [19](#)
- modelo de razão de quadráticos  
na Regressão não linear [19](#)
- modelo de Richards  
na Regressão não linear [19](#)
- modelo de Verhulst  
na Regressão não linear [19](#)
- modelo de Von Bertalanffy  
na Regressão não linear [19](#)
- modelo de Weibull  
na Regressão não linear [19](#)
- modelo modificado por log  
na Regressão não linear [19](#)
- modelos customizados  
na Regressão logística multinomial [6](#)
- Modelos de fragilidade compartilhados paramétricos  
análise [38](#)  
criar gráfico [43](#)  
critérios [39](#)  
estimativa [40](#)  
exportar [43](#)  
imprimir [41](#)  
modelo [39](#)  
Prever [42](#)
- Modelos De Tempo De Falha Acelerada Paramétrica  
análise [31](#)  
criar gráfico [36](#)  
critérios [32](#)  
estimativa [33](#)  
exportar [36](#)  
imprimir [34](#)  
modelo [32](#)  
Prever [35](#)
- modelos dos principais efeitos  
na Regressão logística multinomial [6](#)
- modelos fatoriais completos  
na Regressão logística multinomial [6](#)
- modelos não lineares  
na Regressão não linear [19](#)
- Modelos paramétricos de fragilidade  
variáveis de status de sobrevivência [44](#)

## O

- ordenada na origem  
incluir ou excluir [6](#)

## P

- potência média relativa  
em Análise de probit [11](#)

## Q

- qualidade do ajuste  
na Regressão logística multinomial [7](#)
- Qui-quadrado de Pearson  
para estimar o valor de escala de dispersão [8](#)

- Qui-quadrado de Pearson (*continuação*)  
qualidade do ajuste [7](#)

## R

- R-quadrado de Cox e Snell  
na Regressão logística multinomial [7](#)
- R-quadrado de McFadden  
na Regressão logística multinomial [7](#)
- R-quadrado de Nagelkerke  
na Regressão logística multinomial [7](#)
- razão de verossimilhança  
para estimar o valor de escala de dispersão [8](#)  
qualidade do ajuste [7](#)
- regressão assintótica  
na Regressão não linear [19](#)
- Regressão de Kernel Ridge  
dobras de validação cruzada [30](#)  
exibição [30](#)  
gráficos [30](#)  
Parâmetros [30](#)  
parâmetros da grade [30](#)  
salvar [30](#)
- Regressão de quantil  
critérios [13](#)  
estatísticas [12](#)  
exemplo [12](#)  
exibição [15](#)  
exportar [17](#)  
modelo [14](#)  
salvar [16](#)
- Regressão linear  
estimação ponderada [22](#)  
Regressão por quadrados mínimos de dois estágios [24](#)
- regressão logística [2](#)
- Regressão logística  
binary [1](#)  
coeficientes [2](#)  
configurar regra [3](#)  
contrastes [3](#)  
corte de classificação [5](#)  
covariáveis categóricas [3](#)  
covariáveis de sequência de caracteres [3](#)  
definir regra de seleção [3](#)  
estatísticas [2](#)  
estatísticas de qualidade de ajuste de Hosmer-Lemeshow [5](#)  
estatísticas e gráficos [5](#)  
exemplo [2](#)  
iterações [5](#)  
medidas de influência [4](#)  
métodos de seleção de variáveis [3](#)  
opções de exibição [5](#)  
probabilidade para stepwise [5](#)  
recursos adicionais do comando [5](#)  
resíduos [4](#)  
salvando novas variáveis [4](#)  
termo constante [5](#)  
valores preditos [4](#)
- regressão logística binária [1](#), [2](#)
- Regressão logística multinomial  
categoria de referência [7](#)  
critérios [8](#)  
estatísticas [7](#)

- Regressão logística multinomial (*continuação*)
  - exportando informações de modelo [9](#)
  - modelos [6](#)
  - recursos adicionais do comando [10](#)
  - salvar [9](#)
- Regressão logística SPSS [2](#)
- Regressão não linear
  - algoritmo de Levenberg-Marquardt [21](#)
  - derivados [21](#)
  - estatísticas [18](#)
  - estimativas bootstrap [21](#)
  - exemplo [18](#)
  - função de perda [20](#)
  - interpretação de resultados [21](#)
  - lógica condicional [18](#)
  - métodos de estimação [21](#)
  - modelo segmentado [18](#)
  - modelos comuns não lineares [19](#)
  - Parâmetros [19](#)
  - programação quadrática sequencial [21](#)
  - recursos adicionais do comando [22](#)
  - resíduos [21](#)
  - restrições paramétrica [20](#)
  - salvar novas variáveis [21](#)
  - valores iniciais [19](#)
  - valores preditos [21](#)
- Regressão por quadrados mínimos de dois estágios
  - covariância paramétricas [25](#)
  - estatísticas [24](#)
  - exemplo [24](#)
  - recursos adicionais do comando [25](#)
  - salvando novas variáveis [25](#)
  - variáveis instrumentais [24](#)
- Regressão proibito
  - estatísticas [10](#)
  - exemplo [10](#)
- regressão restrita
  - na Regressão não linear [20](#)
- restrições paramétrica
  - na Regressão não linear [20](#)
- Ridge do kernel
  - alfa [28](#)
  - coef0 [28](#)
  - gama [28](#)
  - grau [28](#)
  - modelo único [28](#)
  - seleção de modelo [28](#)

## S

- seleção forward
  - em Regressão logística [3](#)
- seleção stepwise
  - em Regressão logística [3](#)
  - na Regressão logística multinomial [6](#)
- separação
  - na Regressão logística multinomial [8](#)
- singularidade
  - na Regressão logística multinomial [8](#)
- Sobrevivência AFT
  - Diálogo sobrevivência-Variáveis de categoria [37](#)

## T

- tabelas de classificação
  - na Regressão logística multinomial [7](#)
- Tabelas de mortalidade
  - variáveis de status de sobrevivência [37](#)
- tabelas de probabilidades de célula
  - na Regressão logística multinomial [7](#)
- termo constante
  - Na Regressão Linear [5](#)
- teste de paralelismo
  - em Análise de probit [11](#)

## V

- valor de escala de dispersão
  - na Regressão logística multinomial [8](#)
- valores de ponto de alavanca
  - em Regressão logística [4](#)





