

IBM SPSS Regression 24

IBM

Comunicado

Antes de usar estas informações e o produto suportado por elas, leia as informações nos “Avisos” na página 33.

Informações sobre o produto

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

Índice

Capítulo 1. Escolhendo um Procedimento para Regressão Logística Binária 1

Capítulo 2. Regressão logística. 3

Regra do conjunto de regressão logística	4
Métodos de seleção de variáveis de regressão logística	4
Definir variáveis categóricas de regressão logística.	5
Regressão logística - Salvar novas variáveis	5
Opções de regressão logística	6
Recursos adicionais do comando LOGISTIC REGRESSION	7

Capítulo 3. Regressão logística multinomial 9

Regressão logística multinomial	9
Criar termos	10
Categoria de referência de regressão logística multinomial	10
Estatísticas de regressão logística multinomial	11
Critérios de Regressão logística multinomial	11
Opções de regressão logística multinomial	12
Regressão logística multinomial - Salvar	12
Recursos adicionais do comando NOMREG	13

Capítulo 4. Análise de probitos 15

Definir intervalo de Análise de probito	16
Opções de análise de probit	16
Recursos adicionais do comando PROBIT	17

Capítulo 5. Regressão não linear 19

Lógica condicional (Regressão não linear)	20
Parâmetros de regressão não linear	20

Modelos comuns de regressão não linear	21
Função de perda de regressão não linear	21
Restrições paramétrica de regressão não linear.	22
Regressão não linear - Salvar novas variáveis	22
Opções de regressão não linear	22
Interpretando resultados de regressão não linear	23
Recursos adicionais do comando NLR	23

Capítulo 6. Estimação Ponderada 25

Opções de estimação ponderada	26
Recursos adicionais do comando WLS	26

Capítulo 7. Regressão por quadrados mínimos de dois estágios 27

Opções de regressão de quadrados mínimos de dois estágios.	28
Recursos adicionais do comando 2SLS	28

Capítulo 8. Esquemas de codificação de variável categórica 29

Desvio	29
Simplex.	29
Helmert	30
Diferença	30
Polinomial.	30
Repetido	31
Especial	31
Indicador	32

Avisos 33

Marcas comerciais	35
-----------------------------	----

Índice Remissivo 37

Capítulo 1. 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.

A **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

A **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 preditas e resíduos por subpopulações
- Correção de estimativas de variância para sobredispersão
- Matriz de covariâncias das estimativas paramétrica
- Testes de combinações lineares paramétricas
- Especificação explícita de modelos aninhados
- Ajustar modelos de regressão logística condicionais correspondentes a 1-1 usando variáveis diferenciadas

Nota: 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.

Nota: 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.

Capítulo 2. 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 clique em **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.

As 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 as variáveis independentes são inseridas na análise. Usando métodos diferentes, é possível construir uma variedade de modelos de regressão a partir do mesmo conjunto de variáveis.

- *Inserir*. 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 são baseados no ajuste de um único modelo. Portanto, os valores de significância geralmente são inválidos quando é usado um método stepwise.

Todas as variáveis independentes selecionadas são incluídas em um único modelo de regressão. No entanto, é possível especificar diferentes métodos de entrada para diferentes subconjuntos de variáveis. Por exemplo, é possível inserir um bloco de variáveis no modelo de regressão usando a seleção stepwise e um segundo bloco usando 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.** Contrastos 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 predita da 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 mais alto; por exemplo, se a variável dependente usar os valores 0 e 1, a probabilidade predita da categoria 1 será salva.

- *Associação ao Grupo Predito.* O grupo com a maior probabilidade posterior, com base nos 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 da 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.
- *DfBeta(s).* A diferença no valor beta é a mudança no coeficiente de regressão resultante da exclusão de um caso específico. 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 ele for predito 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 do 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 para arquivo XML. As estimativas paramétrica e (opcionalmente) suas covariâncias são exportadas para o 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 do ajuste é mais robusta que a estatística de qualidade do ajuste tradicional usada em regressão logística, especialmente para modelos com covariáveis e estudos contínuos com tamanhos pequenos de amostras. 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 escore 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.

Capítulo 3. Regressão logística multinomial

A Regressão logística multinomial é útil para situações nas quais você deseja poder classificar assuntos com base em 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 previstas por categoria de resposta. Tabulação cruzada: frequências observadas e previstas (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 > 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 de 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 & 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.

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

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

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

Todos de 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 de 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 previstas.
- **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 usado se o valor for 0. Especifique um valor não negativo.
- **Convergência paramétrica.** 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 especificar um valor não negativo menor que 1. Esse valor é incluído em cada célula vazia da tabulação cruzada da 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 score. 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 estimadas.** 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.

Capítulo 4. Análise de probitos

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. Aplicando a análise de probit a esses dados, é possível determinar a intensidade do relacionamento entre a concentração e a morte, e determinar qual seria a concentração apropriada de pesticida se você quiser ter certeza de que matará, digamos que 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.

Esse procedimento usa algoritmos propostos e implementados em NPSOL[®] por Gill, Murray, Saunders & Wright para estimar os parâmetros de modelo.

Considerações de dados da análise de probit

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 de probit

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 selecionar, clique em **Definir intervalo** para definir 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 da 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 das potências médias 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.* Estima 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 de uma ou mais covariáveis é 0. O probit estima a taxa de resposta natural usando a proporção de respostas para o nível de controle como um valor inicial.
- *Valor.* Configura a taxa de resposta natural no modelo (selecione este item quando você 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.

Capítulo 5. 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 paramétrica e soma de quadrados dos resíduos. Para cada modelo: soma de quadrados para regressão, resíduos, total não corrigido e total corrigido, estimativas paramétrica, erros padrão assintóticos e matriz de correlação assintótica de estimativas paramétrica.

Nota: A regressão não linear restrita usa os algoritmos propostos e implementados em NPSOL[®] por Gill, Murray, Saunders e Wright para estimar os parâmetros de 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 Modelo 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 é igual a 0 para $X \leq 0$, X para $0 < X < 1$ e 1 para $X \geq 1$. 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 de que desigualdades duplas, como $0 < X < 1$, devem ser escritas como expressões compostas, como $(X > 0 \ \& \ X < 1)$.

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	$(b1 + b2 * x) ** (-1 / b3)$
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 \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 Preditos*. Salva valores preditos com o nome de variável pred_.
- *Derivados*. Um derivado 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*. Essa opção estará 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 para estimar o erro padrão de uma estatística utilizando 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 irrestritos. 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*. Esta é o algoritmo padrão para modelos irrestritos. 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. Na caixa de 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 de análise anterior** na caixa de diálogo Parâmetros para continuar a interação 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.

Capítulo 6. Estimação Ponderada

Os modelos de regressão linear padrão consideram que a variância é constante na população em estudo. Quando esse não for o caso -- por exemplo, quando casos que estão na parte superior em algum atributo mostrarem maior variabilidade do que casos que estão na parte inferior nesse atributo -- a regressão linear usando 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 paramétricas 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, tempo de sobrevivência após a cirurgia -- use Tabelas de mortalidade, Kaplan-Meier ou Regressão de Cox, disponível na opção 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 na opção Estatísticas avançadas.

Obtendo uma análise de estimação ponderada

1. Nos menus, escolha:

Analisar > Regressão > Estimação ponderada...

2. Selecione uma variável dependente.
3. Selecionar 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 desta variável elevado 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.
 - *Intervalo de Potências.* Isso é utilizado 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.

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

Capítulo 7. 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 paramétrica.

Considerações de dados de regressão de quadrados mínimos de dois estágios

Dados. As variáveis dependentes e independentes devem ser quantitativas. Variáveis categóricas, como religião, maioria 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 seus dados não forem independentes -- por exemplo, se você observar a mesma pessoa em várias condições -- use o procedimento Medidas repetidas, disponível na opção Modelos avançados.

Obtendo uma análise de regressão de quadrados mínimos de dois estágios

1. Nos menus, escolha:
Analisar > Regressão > Quadrados mínimos de dois estágios...
2. Selecione uma variável dependente.
3. Selecione uma ou mais variáveis explicativas (preditoras).

4. Selecione uma ou mais variáveis instrumentais.

- *Instrumental*. Essas são as variáveis utilizadas para calcular os valores preditos para as variáveis endógenas no primeiro estágio de uma 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 paramétrica.

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.

Capítulo 8. Esquemas de codificação de variável categórica

Em muitos procedimentos, é possível solicitar a substituição automática de uma variável independente categórica com um conjunto de variáveis de contraste, que serão, então, inseridos ou removidos de uma equação como um bloco. É possível especificar como o conjunto de variáveis de contraste deve ser codificado, geralmente, no subcomando CONTRAST. Este apêndice explica e ilustra quão diferentes os tipos de contraste solicitados em CONTRAST realmente funcionam.

Desvio

Desvio da média global. Em termos de matriz, esses contrastes têm o formato:

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

em que k é o número de categorias para a variável independente e a última categoria é omitida por padrão. Por exemplo, os contrastes de desvio para uma variável independente com três categorias são as seguintes:

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

Para omitir uma categoria diferente da última, especifique o número da categoria omitida entre parênteses após a palavra-chave DEVIATION. Por exemplo, o subcomando a seguir obtém os desvios para as categorias primeira e terceira e omite a segunda:

```
/CONTRAST(FACTOR)=DEVIATION(2)
```

Suponha que *fator* tem três categorias. A matriz de contraste resultante será

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

Simples

Contrastes simples. Compara cada nível de um fator com o último. O formulário matriz geral é

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

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

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

Para usar outra categoria em vez da última como uma categoria de referência, especifique entre parênteses após a palavra-chave SIMPLE o número de sequência da categoria de referência, que não é necessariamente o valor associado a essa categoria. Por exemplo, o seguinte subcomando CONTRAST obtém uma matriz de contraste que omite a segunda categoria:

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

Suponha que *fator* tenha quatro categorias. A matriz de contraste resultante será

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

Helmert

Contrastes de Helmert. Compara as categorias de uma variável independente com a média das categorias subsequentes. O formulário matriz geral é

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

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

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

Diferença

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

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

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

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

Polinomial

Contrastes polinominais ortogonais. O primeiro grau de liberdade contém o efeito linear em todas as categorias; o segundo grau de liberdade, o efeito quadrático; o terceiro grau de liberdade, o cúbico, e assim por diante, para o efeito de ordem mais alta.

É possível especificar o espaçamento entre os níveis de tratamento medido pela variável categórica fornecida. O espaçamento igual, que é o padrão, se você omitir a métrica, pode ser especificado como números inteiros consecutivos de 1 a k , em que k é o número de categorias. Se a variável *drug* possuir três categorias, o subcomando

```
/CONTRAST(DRUG)=POLYNOMIAL
```

é o mesmo que

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

O espaçamento igual não é sempre necessário, no entanto. Por exemplo, suponha que *drug* representa dosagens diferentes de um determinado medicamento para os três grupos. Se a dosagem administrada ao segundo grupo for duas vezes maior que a dosagem fornecida ao primeiro grupo, e a dosagem

administrada ao terceiro grupo for três vezes maior que para o primeiro grupo, as categorias de tratamento serão igualmente espaçadas e uma métrica apropriada para esta situação consiste em números inteiros consecutivos:

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

Se, no entanto, a dosagem administrada para o segundo grupo for quatro vezes maior que a fornecida para o primeiro grupo, e a dosagem administrada para o terceiro grupo for sete vezes maior que a fornecida para o primeiro grupo, uma métrica apropriada seria

```
/CONTRAST(DRUG)=POLYNOMIAL(1,4,7)
```

Em ambos os casos, o resultado da especificação de contraste é que o primeiro grau de liberdade para *drug* contém o efeito linear dos níveis de dosagem e o segundo grau de liberdade contém o efeito quadrático.

Contrastes polinomiais são especialmente úteis em testes de tendências e para investigar a natureza das superfícies de resposta. Também é possível usar os contrastes polinomiais para executar o ajuste de curva não linear, como regressão curvilínea.

Repetido

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

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

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

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

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

Especial

Um contraste definido pelo usuário. Permite a entrada de contrastes especiais na forma de matrizes quadradas com tantas linhas e colunas quanto existirem categorias da variável independente fornecida. Para MANOVA e LOGLINEAR, a primeira linha inserida sempre será a média ou constante, efeito e representa o conjunto de ponderações, indicando como obter a média de outras variáveis independentes, se houver, sobre a variável fornecida. Geralmente, esse contraste é um vetor de uns.

As linhas restantes da matriz contêm os contrastes especial indicando as comparações entre as categorias da variável. Geralmente, contrastes ortogonais são os mais úteis. Os contrastes ortogonais são estatisticamente independentes e são não redundantes. Os contrastes são ortogonais se:

- Para cada linha, os coeficientes de contraste somarem 0.
- Os produtos de coeficientes correspondentes para todos os pares de linhas separadas também somarem 0.

Por exemplo, suponha que o tratamento possua quatro níveis e que você deseja comparar os vários níveis de tratamento uns com os outros. Um contraste especial apropriado é

```
(1  1  1  1) weights for mean calculation
(3 -1 -1 -1) compare 1st with 2nd through 4th
(0  2 -1 -1) compare 2nd with 3rd and 4th
(0  0  1 -1) compare 3rd with 4th
```

que é especificado por meio do subcomando CONTRAST a seguir para o MANOVA, LOGISTIC REGRESSION e o COXREG:

```
/CONTRAST(TREATMNT)=SPECIAL( 1 1 1 1
                             3 -1 -1 -1
                             0 2 -1 -1
                             0 0 1 -1 )
```

Para LOGLINEAR, é necessário especificar:

```
/CONTRAST(TREATMNT)=BASIS SPECIAL( 1 1 1 1
                                   3 -1 -1 -1
                                   0 2 -1 -1
                                   0 0 1 -1 )
```

Cada linha excetua as somas de linha média para 0. Os produtos de cada par de linhas separadas somam 0 também:

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

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

Indicador

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

Avisos

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

LICENÇA DE COPYRIGHT:

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

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

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

Marcas comerciais

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

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

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

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

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

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

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

Índice Remissivo

A

Análise de probitos
critérios 16
definir intervalo 16
estatísticas 15, 16
example 15
intervalos de confiança fiduciários 16
iterações 16
potência média relativa 16
recursos adicionais do comando 17
taxa de resposta natural 16
teste de paralelismo 16

C

categoria de referência
na Regressão logística multinomial 10
células com zero observações
na Regressão logística multinomial 11
classificação
na Regressão logística multinomial 9
contrastes
em Regressão logística 5
covariáveis
em Regressão logística 5
covariáveis categóricas 5
covariáveis de sequência de caracteres
em Regressão logística 5
critério de convergência
na Regressão logística multinomial 11

D

delta
como correção para as células com zero observações 11
DfBeta
em Regressão logística 5

E

eliminação backward
em Regressão logística 4
estatísticas de qualidade de ajuste de Hosmer-Lemeshow
em Regressão logística 6
Estimatação Ponderada 25
estatísticas 25
example 25
exibir ANOVA e estimativas 26
histórico de iteração 26
log da verossimilhança 25
recursos adicionais do comando 26
salvar melhores ponderações como nova variável 26

estimativas paramétrica
na Regressão logística multinomial 11
etapa pela metade
na Regressão logística multinomial 11

F

função deviance
para estimar o valor de escala de dispersão 12

H

histórico de iteração
na Regressão logística multinomial 11

I

ID de Cook
em Regressão logística 5
intervalos de confiança
na Regressão logística multinomial 11
intervalos de confiança fiduciários
em Análise de probit 16
iterações
em Análise de probit 16
em Regressão logística 6
na Regressão logística multinomial 11

L

lei dos retornos decrescentes de Metcherlich
na Regressão não linear 21
log da verossimilhança
em Estimatação ponderada 25
na Regressão logística multinomial 11

M

matriz de correlações
na Regressão logística multinomial 11
matriz de covariâncias
na Regressão logística multinomial 11
modelo de densidade
na Regressão não linear 21
modelo de densidade de lucro
na Regressão não linear 21
modelo de Gauss
na Regressão não linear 21

modelo de Gompertz
na Regressão não linear 21
modelo de Johnson-Schumacher
na Regressão não linear 21
modelo de Michaelis-Menten
na Regressão não linear 21
modelo de Morgan-Mercer-Flourin
na Regressão não linear 21
modelo de Peal-Reed
na Regressão não linear 21
modelo de razão de cúbicos
na Regressão não linear 21
modelo de razão de quadráticos
na Regressão não linear 21
modelo de Richards
na Regressão não linear 21
modelo de Verhulst
na Regressão não linear 21
modelo de Von Bertalanffy
na Regressão não linear 21
modelo de Weibull
na Regressão não linear 21
modelo modificado por log
na Regressão não linear 21
modelos customizados
na Regressão logística multinomial 9
modelos dos principais efeitos
na Regressão logística multinomial 9
modelos fatoriais completos
na Regressão logística multinomial 9
modelos não lineares
na Regressão não linear 21

O

ordenada na origem
incluir ou excluir 9

P

potência média relativa
em Análise de probit 16

Q

qualidade do ajuste
na Regressão logística multinomial 11
Qui-quadrado de Pearson
para estimar o valor de escala de dispersão 12
qualidade do ajuste 11

R

R-quadrado de Cox e Snell
na Regressão logística multinomial 11

- R-quadrado de McFadden
 - na Regressão logística multinomial 11
 - R-quadrado de Nagelkerke
 - na Regressão logística multinomial 11
 - razão de verossimilhança
 - para estimar o valor de escala de dispersão 12
 - qualidade do ajuste 11
 - regressão assintótica
 - na Regressão não linear 21
 - Regressão linear
 - estimação ponderada 25
 - Regressão por quadrados mínimos de dois estágios 27
 - Regressão logística 3
 - binary 1
 - coeficientes 3
 - contrastes 5
 - corte de classificação 6
 - covariáveis categóricas 5
 - covariáveis de sequência de caracteres 5
 - definir regra de seleção 4
 - estatísticas 3
 - estatísticas de qualidade de ajuste de Hosmer-Lemeshow 6
 - estatísticas e gráficos 6
 - example 3
 - iterações 6
 - medidas de influência 5
 - métodos de seleção de variáveis 4
 - opções de exibição 6
 - probabilidade para stepwise 6
 - recursos adicionais do comando 7
 - regra do conjunto 4
 - resíduos 5
 - salvando novas variáveis 5
 - termo constante 6
 - valores preditos 5
 - regressão logística binária 1
 - Regressão logística multinomial 9, 11
 - categoria de referência 10
 - critérios 11
 - estatísticas 11
 - exportando informações de modelo 12
 - modelos 9
 - recursos adicionais do comando 13
 - salvar 12
 - Regressão não linear 19
 - algoritmo de Levenberg-Marquardt 22
 - derivados 22
 - estatísticas 19
 - estimativas bootstrap 22
 - example 19
 - função de perda 21
 - interpretação de resultados 23
 - lógica condicional 20
 - métodos de estimação 22
 - modelo segmentado 20
 - modelos comuns não lineares 21
 - parâmetros 20
 - programação quadrática sequencial 22
 - Regressão não linear (*continuação*)
 - recursos adicionais do comando 23
 - resíduos 22
 - restrições paramétrica 22
 - salvar novas variáveis 22
 - valores iniciais 20
 - valores preditos 22
 - Regressão por quadrados mínimos de dois estágios 27
 - covariância paramétricas 28
 - estatísticas 27
 - example 27
 - recursos adicionais do comando 28
 - salvando novas variáveis 28
 - variáveis instrumentais 27
 - regressão restrita
 - na Regressão não linear 22
 - restrições paramétrica
 - na Regressão não linear 22
- S**
- seleção forward
 - em Regressão logística 4
 - seleção stepwise
 - em Regressão logística 4
 - na Regressão logística multinomial 9
 - separação
 - na Regressão logística multinomial 11
 - singularidade
 - na Regressão logística multinomial 11
- T**
- tabelas de classificação
 - na Regressão logística multinomial 11
 - tabelas de probabilidades de célula
 - na Regressão logística multinomial 11
 - termo constante
 - na regressão linear 6
 - teste de paralelismo
 - em Análise de probit 16
- V**
- valor de escala de dispersão
 - na Regressão logística multinomial 12
 - valores de ponto de alavanca
 - em Regressão logística 5



Impresso no Brasil