

IBM SPSS Complex Samples 24

IBM

Comunicado

Antes de usar estas informações e o produto suportado por elas, leia as informações nos “Avisos” na página 59.

Informações sobre o produto

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

Índice

Capítulo 1. Introdução a procedimentos de amostras complexas 1

Propriedades das amostras complexas	1
Uso de procedimentos de Amostras Complexas.	2
Arquivos de plano	2
Leituras adicionais	3

Capítulo 2. Amostragem de um design complexo 5

Criando um novo plano de amostra	5
Assistente de amostragem: Variáveis de design	5
Controles de árvore para navegação no Assistente de amostragem	6
Assistente de amostragem: Método de amostragem	6
Assistente de amostragem: Tamanho da amostra	7
Definir tamanhos desiguais	8
Assistente de amostragem: Variáveis de saída	8
Assistente de amostragem: Sumarização do plano	9
Assistente de amostragem: Exibir opções de seleção de amostra	9
Assistente de amostragem: Exibir arquivos de saída de amostra	9
Assistente de amostragem: Concluir	10
Modificando um plano de amostra existente	10
Assistente de amostragem: Sumarização do plano	10
Executando um plano de amostra existente.	10
Recursos adicionais de comandos CSPLAN e CSSELECT.	11

Capítulo 3. Preparando uma Amostra Complexa para análise 13

Criando um novo plano de análise	13
Assistente de preparação de análise: variáveis de design	13
Controles de árvore para navegação no Assistente de análise	14
Assistente de preparação de análise: método de estimação	14
Assistente de preparação de análise: tamanho	14
Definir tamanhos desiguais	15
Análise do assistente de preparação: resumo do plano	15
Assistente de preparação de análise: conclusão	15
Modificando um plano de análise existente.	16
Análise do assistente de preparação: resumo do plano	16

Capítulo 4. Plano de amostras complexo 17

Capítulo 5. Frequências de amostras complexas 19

Estatísticas de frequências de amostras complexas	19
Valores omissos de amostras complexas	20

Opções de amostras complexas.	20
---------------------------------------	----

Capítulo 6. Descrições de amostras complexas 23

Estatísticas descritivas de amostras complexas.	23
Valores omissos de amostras complexas descritivas	24
Opções de amostras complexas.	24

Capítulo 7. Tabulações cruzadas de amostras complexas. 25

Estatísticas de Tabulações cruzadas de amostras complexas.	25
Valores omissos de amostras complexas	26
Opções de amostras complexas.	27

Capítulo 8. Razões de amostras complexas 29

Estatísticas de razão de amostras complexas	29
Valores omissos de razões de amostras complexas	30
Opções de amostras complexas.	30

Capítulo 9. Modelo linear geral de amostras complexas. 31

Modelo linear geral de amostras complexas	31
Estatísticas de modelo linear geral de amostras complexas	32
Testes de hipótese de amostras complexas	33
Médias estimadas do modelo linear geral de amostras complexas	33
Modelo linear geral de Amostras complexas - Salvar	34
Opções de modelo linear geral de amostras complexas	34
Recursos adicionais de comando CSGLM	35

Capítulo 10. Regressão logística de amostras complexas. 37

Categoria de referência de regressão logística de amostras complexas	37
Modelo de regressão logística de Amostras Complexas	38
Estatísticas de Regressão logística de amostras complexas	38
Testes de hipótese de amostras complexas	39
Razões de chances de regressão logística de amostras complexas	40
Regressão logística de amostras complexas - Salvar	40
Opções de Regressão logística de amostras complexas	40
Recursos adicionais do comando CSLOGISTIC	41

Capítulo 11. Regressão ordinal de amostras complexas. 43

Probabilidades de resposta de Regressão ordinal de amostras complexas 44
Modelo de regressão ordinal de amostras complexas 44
Estatísticas de Regressão ordinal de Amostras complexas 44
Testes de hipótese de amostras complexas 45
Razões de chances de regressão ordinal de amostras complexas 46
Regressão ordinal de amostras complexas - Salvar 46
Opções de regressão ordinal de amostras complexas 47
Recursos adicionais de comando CSORDINAL . . . 47

Capítulo 12. Regressão de Cox de amostras complexas. 49

Definir evento 50

Preditores 51
 Definir preditor dependente de tempo 51
Subgrupos. 51
Modelo. 52
Estatísticas. 52
Gráficos 53
Testes de hipótese 54
Salvar o 54
Exportar 55
Opções de. 56
Recursos adicionais de comando CSCOXREG . . . 57

Avisos 59

Marcas comerciais 61

Índice Remissivo 63

Capítulo 1. Introdução a procedimentos de amostras complexas

Uma suposição inerente de procedimentos analíticos em pacotes de software tradicionais é que as observações em um arquivo de dados representam uma amostra aleatória simples da população de interesse. Essa suposição é insustentável para um número crescente de empresas e pesquisadores que acham que ela tem um custo reduzido e é conveniente para obter amostras de uma maneira mais estruturada.

A opção Amostras complexas permite selecionar uma amostra de acordo com um design complexo e incorporar as especificações de design na análise de dados, assegurando, portanto, que seus resultados sejam válidos.

Propriedades das amostras complexas

Uma amostra complexa pode diferir de uma amostra aleatória simples de várias maneiras. Em uma amostra aleatória simples, unidades individuais de amostragem são selecionadas aleatoriamente com probabilidade igual e sem substituição (WOR) diretamente da população inteira. Por outro lado, uma determinada amostra complexa pode ter alguns ou todos os seguintes recursos:

Estratificação. A amostragem estratificada envolve selecionar amostras independentemente em subgrupos sem sobreposição da população ou camadas. Por exemplo, camadas podem ser grupos socioeconômicos, categorias de cargos, grupos de idade ou grupos étnicos. Com a estratificação, é possível assegurar tamanhos de amostra adequados para os subgrupos de interesse, melhorar a precisão de estimativas gerais e usar métodos de amostragem diferentes de camada para camada.

Armazenamento em cluster. A amostragem de cluster envolve a seleção de grupos de unidades de amostragem ou clusters. Por exemplo, os clusters podem ser escolas, hospitais ou áreas geográficas, e as unidades de amostragem podem ser estudantes, pacientes ou cidadãos. O clusterização é comum em designs de estágios múltiplos e amostras de área (geográfica).

Estágios múltiplos. Na amostragem de estágios múltiplos, selecione uma amostra de primeiro estágio com base em clusters. Em seguida, crie uma amostra de segundo estágio, colhendo subamostras a partir dos clusters selecionados. Se a amostra de segundo estágio for baseada em subclusters, será possível, então, incluir um terceiro estágio na amostra. Por exemplo, no primeiro estágio de uma pesquisa, uma amostra de cidades pode ser colhida. Em seguida, a partir das cidades selecionadas, as famílias podem ser amostradas. Por último, a partir da família selecionada, indivíduos podem ser pesquisados. Os assistentes Amostragem e Preparação de Análise permitem especificar três estágios em um design.

Amostragem não aleatória. Quando a seleção aleatória é difícil de obter, a amostra das unidades pode ser obtida sistematicamente (em um intervalo fixo) ou sequencialmente.

Probabilidades de seleção diferentes. Quando houver clusters de amostragem que contêm números diferentes de unidades, será possível usar a amostragem de probabilidade proporcional ao tamanho (PPS) para tornar a probabilidade de seleção de um cluster igual à proporção de unidades que ele contém. As amostragens de PPS também podem usar esquemas de ponderação mais gerais para selecionar unidades.

Amostragem irrestrita. A amostragem irrestrita seleciona unidades com substituição (WR). Assim, uma unidade individual pode ser selecionada para a amostra mais de uma vez.

Ponderações de amostragem. As ponderações de amostragem são calculadas automaticamente ao colher uma amostra complexa e idealmente correspondem à "frequência" que cada unidade de amostragem

representa na população alvo. Portanto, a soma das ponderações sobre a amostra deve estimar o tamanho da população. Os procedimentos de análise de amostras complexas requerem ponderações de amostragem para analisar corretamente uma amostra complexa. Observe que essas ponderações devem ser usadas inteiramente dentro da opção Amostras Complexas e não devem ser usadas com outros procedimentos analíticos por meio do procedimento Casos de Ponderação, que trata das ponderações como replicações de caso.

Uso de procedimentos de Amostras Complexas

O uso de procedimentos de Amostras Complexas depende de suas necessidades específicas. Os principais tipos de usuários são aqueles que:

- Planejam e realizam pesquisas de acordo com os designs complexos, possivelmente analisando a amostra num momento posterior. A principal ferramenta para os pesquisadores é o Assistente de amostragem.
- Analisam os arquivos de dados de amostra anteriormente obtidos de acordo com os designs complexos. Antes de usar os procedimentos de análise de Amostras Complexas, pode ser necessário usar o Assistente de preparação de análise.

Independentemente de qual tipo de usuário você é, é necessário fornecer informações de design para procedimentos de Amostras Complexas. Essas informações são armazenadas em um **arquivo de plano** para facilitar a reutilização.

Arquivos de plano

Um arquivo de plano contém especificações de amostra complexas. Há dois tipos de arquivos de plano:

Plano de amostragem. As especificações fornecidas no Assistente de amostragem definem um plano de amostra que é usado para colher uma amostra complexa. O arquivo de plano de amostragem contém essas especificações. O arquivo de plano de amostragem também contém um plano de análise padrão que usa os métodos de estimação adequados para o plano de amostra especificado.

Plano de análise. Este arquivo de plano contém informações necessárias pelos procedimentos de análise de Amostras Complexas para calcular adequadamente estimativas de variação para uma amostra complexa. O plano inclui a estrutura de amostra, métodos de estimação para cada estágio e referências a variáveis necessárias, como SampleWeight. O Assistente de preparação de análise permite criar e editar planos de análise.

Existem várias vantagens em salvar suas especificações em um arquivo de plano, incluindo:

- Um pesquisador pode especificar o primeiro estágio de um plano de amostragem de vários estágios e colher as unidades de primeiro estágio agora, coletar informações sobre unidades de amostragem para o segundo estágio e, então, modificar o plano de amostragem para incluir o segundo estágio.
- Um analista que não tem acesso ao arquivo de plano de amostragem pode especificar um plano de análise e referir-se a esse plano a partir de cada procedimento de análise de Amostras Complexas.
- Um designer de amostras de uso público de grande escala pode publicar o arquivo de plano de amostragem, que simplifica as instruções para analistas e evita a necessidade de cada analista especificar seus próprios planos de análise.

Leituras adicionais

Para obter mais informações sobre técnicas de amostragem, consulte os seguintes textos:

Cochran, W. G. 1977. *Sampling Techniques*, 3ª ed. Nova York: John Wiley and Sons.

Kish, L. 1965. *Survey Sampling*. Nova York: John Wiley and Sons.

Kish, L. 1987. *Statistical Design for Research*. Nova York: John Wiley and Sons.

Murthy, M. N. 1967. *Sampling Theory and Methods*. Calcutá, Índia: Statistical Publishing Society.

Särndal, C., B. Swensson, e J. Wretman. 1992. *Model Assisted Survey Sampling*. Nova York: Springer-Verlag.

Capítulo 2. Amostragem de um design complexo

O Assistente de amostragem o orienta pelos passos para criar, modificar ou executar um arquivo de plano de amostragem. Antes de usar o Assistente, você deve ter em mente uma população alvo bem definida, uma lista de unidades de amostragem e um plano de amostra apropriado.

Criando um novo plano de amostra

1. Nos menus, escolha:
Analisar > Amostras complexas > Selecionar uma amostra...
2. Selecione **Projetar uma amostra** e escolha um nome de arquivo de plano para salvar o plano de amostra.
3. Clique em **Avançar** para continuar através do assistente.
4. Opcionalmente, no passo Variáveis de design, é possível definir estrato, clusters e ponderações de amostra de entrada. Após defini-los, clique em **Avançar**.
5. Opcionalmente, no passo Método de amostragem, é possível escolher um método para selecionar itens.
Se você selecionar **PPS Brewer** ou **PPS Murthy**, será possível clicar em **Concluir** para exibir a amostra. Caso contrário, clique em **Avançar** e, em seguida:
6. No passo Tamanho da amostra, especifique o número ou a proporção de unidades para amostragem.
7. Agora é possível clicar em **Concluir** para exibir a amostra.

Opcionalmente, em passos adicionais, é possível:

- Escolher variáveis de saída para salvar.
- Incluir um segundo ou terceiro estágio no design.
- Configurar várias opções de seleção, incluindo de quais estágios exibir as amostra, a valor semente de número aleatório e tratar ou não valores omissos de usuário como valores válidos de variáveis de design.
- Escolher onde salvar os dados de saída.
- Colar suas seleções como sintaxe de comando.

Assistente de amostragem: Variáveis de design

Esse passo permite selecionar variáveis de estratificação e de clusterização e definir ponderações de amostra de entrada. Também é possível especificar um rótulo para o estágio.

Estratificar por. A classificação cruzada de variáveis de estratificação define subpopulações ou camadas distintos. Amostras separadas são obtidas para cada camada. Para melhorar a precisão de suas estimativas, as unidades no camada devem ser o mais homogêneas possível para as características de interesse.

Clusters. As variáveis de cluster definem grupos de unidades observacionais ou clusters. Os clusters são úteis quando a amostragem direta de unidades observacionais da população é dispendiosa ou impossível; em vez disso, é possível fazer a amostragem de clusters a partir da população e, em seguida, fazer a amostragem de unidades observacionais dos clusters selecionados. No entanto, o uso de clusters pode introduzir correlações entre unidades de amostragem, resultando em uma perda de precisão. Para minimizar esse efeito, as unidades em clusters devem ser o mais heterogêneas possível para as características de interesse. Deve-se definir pelo menos uma variável de cluster para planejar um design

de vários estágios. Os clusters também são necessários no uso de vários métodos de amostragem diferentes. Consulte o tópico “Assistente de amostragem: Método de amostragem” para obter mais informações

Ponderação de amostra de entrada. Se o plano de amostra atual for parte de um plano de amostra maior, você tem ponderações de amostra de um estágio anterior do design maior. É possível especificar uma variável numérica contendo essas ponderações no primeiro estágio do design atual. As ponderações de amostra são calculadas automaticamente para estágios subsequentes do design atual.

Rótulo de estágio. É possível especificar um rótulo de sequência de caracteres opcional para cada estágio. Isso é usado na saída para ajudar a identificar informações de stagewise.

Nota: A lista de variáveis de origem tem o mesmo conteúdo nos passos do Assistente. Em outras palavras, as variáveis removidas da lista de origem em um determinado passo são removidas da lista em todos os passos. As variáveis retornadas à lista de origem aparecem na lista em todos os passos.

Controles de árvore para navegação no Assistente de amostragem

No lado esquerdo de cada etapa do Assistente de amostragem há uma estrutura de tópicos de todas as etapas. É possível navegar no Assistente clicando no nome de uma etapa ativada na estrutura de tópicos. As etapas são ativadas contanto que todas as etapas anteriores sejam válidas—isto é, contanto que cada etapa anterior tenha fornecido as especificações mínimas obrigatórias para essa etapa. Consulte a Ajuda para etapas individuais para obter mais informações sobre porque uma determinada etapa pode ser inválida.

Assistente de amostragem: Método de amostragem

Esse passo permite especificar como selecionar casos do conjunto de dados ativo.

Método. Os controles nesse grupo são usados para escolher um método de seleção. Alguns tipos de amostragem permitem escolher se fazer a amostragem com substituição (WR) ou sem substituição (WOR). Consulte as descrições de tipo para obter informações adicionais. Observe que alguns tipos de probabilidade-proporcional-ao-tamanho (PPS) estão disponíveis apenas quando os clusters tiverem sido definidos e se todos os tipos de PPS estiverem disponíveis apenas no primeiro estágio de um design. Além disso, os métodos WR estão disponíveis apenas no último estágio de um design.

- **Amostragem aleatória simples.** As unidades são selecionadas com probabilidade igual. Elas podem ser selecionadas com ou sem substituição.
- **Sistemático simples.** As unidades são selecionadas em um intervalo fixo em todo o quadro de amostragem (ou camadas, se tiverem sido especificados) e extraídas sem substituição. Uma unidade selecionada aleatoriamente no primeiro intervalo é escolhida como o ponto de início.
- **Sequencial simples.** As unidades são selecionadas sequencialmente com probabilidade igual e sem substituição.
- **PPS.** Esse método é um método de primeiro estágio que seleciona unidades aleatoriamente com probabilidade proporcional ao tamanho. As unidades podem ser selecionadas com substituição; apenas clusters podem ser amostrados sem substituição.
- **PPS sistemático.** Esse é um método de primeiro estágio que seleciona unidades sistematicamente com probabilidade proporcional ao tamanho. Elas são selecionadas sem substituição.
- **PPS sequencial.** Esse é um método de primeiro estágio que seleciona unidades sequencialmente com probabilidade proporcional ao tamanho do cluster e sem substituição.
- **PPS Brewer.** Esse é um método de primeiro estágio que seleciona dois clusters de cada camada com probabilidade proporcional ao tamanho do cluster e sem substituição. Uma variável de cluster deve ser especificada para usar este método.

- **PPS Murthy.** Esse é um método de primeiro estágio que seleciona dois clusters de cada camada com probabilidade proporcional ao tamanho do cluster e sem substituição. Uma variável de cluster deve ser especificada para usar este método.
- **PPS Sampford.** Esse é um método de primeiro estágio que seleciona mais de dois clusters de cada camada com probabilidade proporcional ao tamanho do cluster e sem substituição. É uma extensão do método de Brewer. Uma variável de cluster deve ser especificada para usar este método.
- **Usar estimação WR para análise.** Por padrão, o método de estimação é especificado no arquivo de plano que está consistente com o método de amostragem selecionado. Isso permite usar a estimação com substituição mesmo que o método de amostragem implique estimação WOR. Esta opção está disponível apenas no estágio 1.

Medida de tamanho (MOS). Se um método PPS for selecionado, deve-se especificar uma medida de tamanho que defina o tamanho de cada unidade. Esses tamanhos podem ser definidos explicitamente em uma variável ou podem ser calculados a partir dos dados. Opcionalmente, é possível configurar limites inferior e superior no MOS, substituindo todos os valores encontrados na variável MOS ou calculados a partir dos dados. Essas opções estão disponíveis apenas no estágio 1.

Assistente de amostragem: Tamanho da amostra

Esse passo permite especificar o número ou proporção de unidades para amostra no estágio atual. O tamanho da amostra pode ser fixo ou pode variar entre os camadas. Para o propósito de especificar o tamanho da amostra, os clusters escolhidos em estágios anteriores podem ser usados para definir o camada.

Unidades. É possível especificar um tamanho exato da amostra ou uma proporção de unidades para amostra.

- **Valor.** Um único valor é aplicado a todos os camadas. Se **Contagens** for selecionado como a métrica de unidade, deve-se inserir um número inteiro positivo. Se **Proporções** for selecionado, deve-se inserir um valor não negativo. A menos que a amostragem seja feita com substituição, os valores de proporção também não devem ser maiores que 1.
- **Valores desiguais para camadas.** Permite inserir valores de tamanho em uma base por estrato por meio da caixa de diálogo Definir valores desiguais.
- **Valores de leitura da variável.** Permite selecionar uma variável numérica que contém valores de tamanho para camadas.

Se **Proporções** for selecionado, haverá a opção de configurar limites inferior e superior no número de unidades amostradas.

Definir tamanhos desiguais

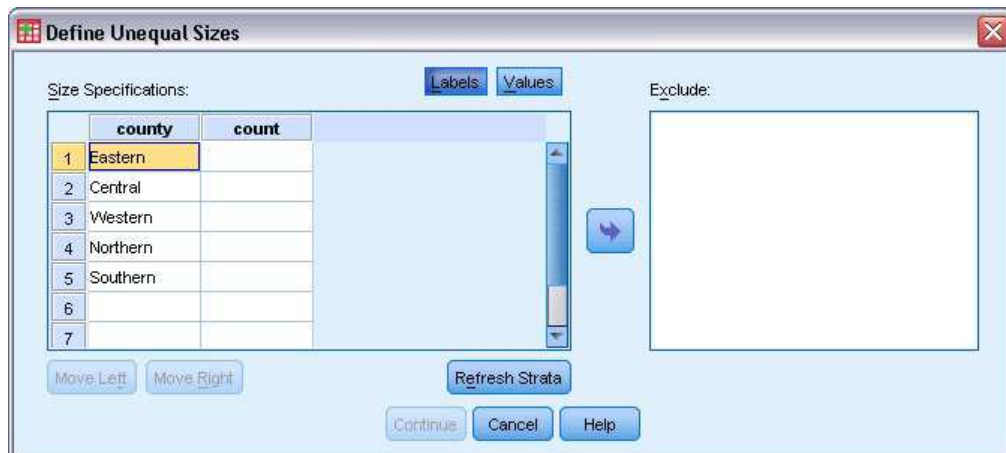


Figura 1. Caixa de diálogo Definir tamanhos desiguais

A caixa de diálogo Definir tamanhos desiguais permite inserir tamanhos em uma base por camada.

Grade de especificações de tamanho. A grade exibe as classificações cruzadas de até cinco variáveis de camada ou cluster — uma combinação de camada/cluster por linha. As variáveis de grade elegíveis incluem todas as variáveis de estratificação dos estágios atual e anterior e todas as variáveis de cluster de estágios anteriores. As variáveis podem ser reordenadas na grade ou movidas para a lista Exclusão. Insira tamanhos na coluna mais à direita. Clique em **Rótulos** ou **Valores** para alternar a exibição de rótulos de valor e de valores de dados para variáveis de estratificação e de cluster nas células de grade. As células que contêm valores não rotulados sempre mostram valores. Clique em **Atualizar camada** para preencher novamente a grade com cada combinação de valores de dados rotulados para variáveis na grade.

Exclusão. Para especificar tamanhos para um subconjunto de combinações de camada/cluster, mova uma ou mais variáveis para a lista Exclusão. Essas variáveis não são usadas para definir tamanhos de amostra.

Assistente de amostragem: Variáveis de saída

Esse passo permite escolher variáveis para salvar quando a amostra é exibida.

Tamanho da população. O número estimado de unidades na população para um determinado estágio. O nome raiz para a variável salva é *PopulationSize_*.

Proporção de amostra. A taxa de amostragem em um determinado estágio. O nome raiz para a variável salva é *SamplingRate_*.

Tamanho da amostra. O número de unidades exibidas em um determinado estágio. O nome raiz para a variável salva é *SampleSize_*.

Ponderação da amostra. O inverso das probabilidades de inclusão. O nome raiz para a variável salva é *SampleWeight_*.

Algumas variáveis stagewise são geradas automaticamente. Eles incluem:

Probabilidades de inclusão. A proporção de unidades exibidas em um determinado estágio. O nome raiz para a variável salva é *InclusionProbability_*.

Ponderação acumulativa. A ponderação da amostra acumulativa durante estágios anteriores ao e incluindo o atual. O nome raiz para a variável salva é *SampleWeightCumulative_*.

Índice. Identifica unidades selecionadas várias vezes em um determinado estágio. O nome raiz para a variável salva é *Index_*.

Nota: Os nomes raiz de variáveis salvas incluem um sufixo de número inteiro que reflete o número do estágio — por exemplo, *PopulationSize_1_* para o tamanho da população salvo para o estágio 1.

Assistente de amostragem: Sumarização do plano

Esse é o último passo em cada estágio, fornecendo uma sumarização das especificações de design de amostra por meio do estágio atual. Aqui, é possível continuar com o próximo estágio (criando-o, se necessário) ou configurar opções para exibir a amostra.

Assistente de amostragem: Exibir opções de seleção de amostra

Esse passo permite escolher se exibir uma amostra. Também é possível controlar outras opções de amostragem, como a semente aleatória e o tratamento de valor omissos.

Desenhar amostra. Além de escolher se exibir uma amostra, também é possível optar por executar parte do plano de amostragem. Os estágios devem ser exibidos por ordem — ou seja, o estágio 2 não pode ser exibido, a menos que o estágio 1 também seja exibido. Ao editar ou executar um plano, não é possível realizar uma nova amostragem de estágios bloqueados.

Semente. Isso permite escolher um valor semente para geração de número aleatório.

Incluir valores omissos de usuário. Isso determina se valores omissos de usuário são válidos. Nesse caso, os valores omissos de usuário são tratados como uma categoria separada.

Dados já ordenados. Se seu quadro de amostra for pré-ordenado pelos valores das variáveis de estratificação, essa opção permitirá acelerar o processo de seleção.

Assistente de amostragem: Exibir arquivos de saída de amostra

Esse passo permite escolher para onde direcionar casos amostrados, variáveis de ponderação, probabilidades conjuntas e regras de seleção de caso.

Dados de amostra. Essas opções permitem determinar onde a saída de amostra será gravada. Ela pode ser incluída no conjunto de dados ativo, gravada em um novo conjunto de dados ou salva em um arquivo de dados externo do IBM® SPSS Statistics. Os conjuntos de dados estão disponíveis durante a sessão atual, mas não estarão disponíveis em sessões subsequentes, a menos que você salve-os explicitamente como arquivos de dados. Os nomes de conjuntos de dados devem estar em conformidade com as regras de nomenclatura de variáveis. Se um arquivo externo ou um novo conjunto de dados for especificado, serão gravadas as variáveis de saída de amostragem e as variáveis no conjunto de dados ativo para os casos selecionados.

Probabilidades conjuntas. Essas opções permitem determinar onde as probabilidades conjuntas serão gravadas. Elas são salvas em um arquivo de dados externo do IBM SPSS Statistics. As probabilidades conjuntas serão produzidas se o método PPS WOR, PPS Brewer, PPS Sampford ou PPS Murthy for selecionado e uma estimação WR não for especificada.

Regras de seleção de caso. Se estiver construindo sua amostra em um estágio por vez, talvez você queira salvar as regras de seleção de caso em um arquivo de texto. Elas são úteis para construir o subquadro para estágios subsequentes.

Assistente de amostragem: Concluir

Este é o passo final. É possível salvar o arquivo de plano e exibir a amostra agora ou colar suas seleções em uma janela de sintaxe.

Ao fazer mudanças em estágios no arquivo de plano existente, é possível salvar o plano editado em um novo arquivo ou sobrescrever o arquivo existente. Ao incluir estágios sem fazer mudanças em estágios existentes, o Assistente sobrescreve automaticamente o arquivo de plano existente. Se desejar salvar o plano em um novo arquivo, selecione **Colar a sintaxe gerada pelo Assistente em uma janela de sintaxe** e mude o nome do arquivo nos comandos de sintaxe.

Modificando um plano de amostra existente

1. Nos menus, escolha:
Analisar > Amostras complexas > Selecionar uma amostra...
2. Selecione **Editar um plano de amostra** e escolha um arquivo de plano para editar.
3. Clique em **Avançar** para continuar através do assistente.
4. Revise o plano de amostragem no passo Sumarização do plano e, em seguida, clique em **Avançar**.
Os passos subsequentes são basicamente os mesmos de um novo design. Consulte a Ajuda para passos individuais para obter informações adicionais.
5. Navegue para o passo Concluir e especifique um novo nome para o arquivo de plano editado ou escolha sobrescrever o arquivo de plano existente.

Como opção, você pode:

- Especifique os estágios que já foram amostrados.
- Remover estágios do plano.

Assistente de amostragem: Sumarização do plano

Este passo permite revisar o plano de amostragem e indicar estágios que já foram amostrados. Se estiver editando um plano, também é possível remover estágios do plano.

Estágios amostrados anteriormente. Se um quadro de amostragem estendido não estiver disponível, será necessário executar um design de amostragem de vários estágios de uma vez. Selecione na lista suspensa quais estágios já foram amostrados. Os estágios que foram executados estão bloqueados; eles não estão disponíveis no passo Exibir opções de seleção de amostra e não podem ser alterados durante a edição de um plano.

Remover estágios. É possível remover os estágios 2 e 3 de um design de vários estágios.

Executando um plano de amostra existente

1. Nos menus, escolha:
Analisar > Amostras complexas > Selecionar uma amostra...
2. Selecione **Exibir uma amostra** e escolha um arquivo de plano para execução.
3. Clique em **Avançar** para continuar através do assistente.
4. Revise o plano de amostragem no passo Sumarização do plano e, em seguida, clique em **Avançar**.
5. Os passos individuais que contêm informações de estágio são ignorados durante a execução de um plano de amostra. Agora é possível acessar o passo Concluir a qualquer momento.

Opcionalmente, é possível especificar estágios que já foram amostrados.

Recursos adicionais de comandos CSPLAN e CSSELECT

O idioma da sintaxe de comando também permite:

- Especifique nomes customizados para variáveis de saída.
- Controle a saída no Visualizador. Por exemplo, é possível suprimir a sumarização stagewise do plano que será exibida se uma mostra for projetada ou modificada, suprimir a sumarização da distribuição de casos amostrados por estrato que serão mostrados se o plano de amostra for executado, e solicitar uma sumarização de processamento de caso.
- Escolha um subconjunto de variáveis no conjunto de dados ativo para gravar em um arquivo de amostra externo ou em um conjunto de dados diferente.

Consulte a *Referência de Sintaxe de Comando* para obter informações de sintaxe completa.

Capítulo 3. Preparando uma Amostra Complexa para análise

O Assistente de Preparação de Análise guia o usuário ao longo das etapas para criar ou modificar um plano de análise para uso com os vários procedimentos de análise de Amostras Complexas. Antes de usar o Assistente, deve-se ter uma amostra colhida, de acordo com um design complexo.

A criação de um novo plano é mais útil quando o usuário não tem acesso ao arquivo do plano de amostragem usado para colher a amostra (lembre-se de que o plano de amostragem contém um plano de análise padrão). Se você não tiver acesso ao arquivo do plano de amostragem usado para colher a amostra, será possível usar o plano de análise padrão contido no arquivo de plano de amostragem ou substituir as especificações de análise padrão e salvar as mudanças em um novo arquivo.

Criando um novo plano de análise

1. Nos menus, escolha:
Analisar > Amostras complexas > Preparar para análise...
2. Selecione **Criar um arquivo de plano** e escolha um nome de arquivo de plano para o qual o plano de análise será salvo.
3. Clique em **Avançar** para continuar pelo assistente.
4. Especifique a variável que contém as ponderações de amostra na etapa Variáveis de Design, definindo opcionalmente camadas e clusters.
5. É possível agora clicar em **Concluir** para salvar o plano.

Opcionalmente, em etapas adicionais é possível:

- Selecionar o método para estimar os erros padrão na etapa Método de Estimação.
- Especificar o número de unidades amostradas ou a probabilidade de inclusão por unidade na etapa Tamanho.
- Incluir um segundo ou terceiro estágio para o design.
- Colar seleções como sintaxe de comando.

Assistente de preparação de análise: variáveis de design

Esta etapa permite identificar as variáveis de estratificação e clusterização e definir ponderações da amostra. Também é possível fornecer um rótulo para o estágio.

camada. A classificação cruzada das variáveis de estratificação define subpopulações distintas ou camada. O total de amostra representa a combinação de amostras independentes a partir de cada camada.

Clusters. As variáveis de cluster definem grupos de unidades observacionais, ou clusters. As amostras colhidas em vários estágios selecionam clusters nos estágios anteriores e, em seguida, unidades de subamostra a partir dos clusters selecionados. Ao analisar um arquivo de dados obtido por uma amostragem de clusters com substituição, você deverá incluir o índice de duplicação como uma variável de cluster.

Ponderação de amostra. Deve-se fornecer ponderações de amostra no primeiro estágio. As ponderações de amostra são calculadas automaticamente para os estágios subsequentes do design atual.

Rótulo de estágio. É possível especificar um rótulo de sequência opcional para cada estágio. Isso é usado na saída para ajudar a identificar informações stagewise.

Nota: a lista de variáveis de origem possui o mesmo conteúdo em etapas do assistente. Em outras palavras, as variáveis removidas da lista de origem em uma etapa específica são removidas da lista em todas as etapas. As variáveis retornadas para a lista de origem são mostradas em todas as etapas.

Controles de árvore para navegação no Assistente de análise

No lado esquerdo de cada etapa do Assistente de análise há uma estrutura de tópicos de todas as etapas. É possível navegar no Assistente clicando no nome de uma etapa ativada na estrutura de tópicos. As etapas são ativadas contanto que todas as etapas anteriores sejam válidas—isto é, contanto que cada etapa anterior tenha fornecido as especificações mínimas obrigatórias para essa etapa. Para obter mais informações sobre porque a etapa pode ser inválida, consulte a Ajuda para etapas individuais.

Assistente de preparação de análise: método de estimação

Esta etapa permite especificar um método de estimação para o estágio.

WR (amostragem com substituição). A estimação WR não inclui uma correção para a amostragem a partir de uma população finita (FPC) ao estimar a variação sob o plano de amostragem complexo. É possível optar por incluir ou excluir o FPC ao estimar a variação sob amostragem aleatória simples (SRS).

Optar por não incluir o FPC na estimação de variação SRS é recomendado quando as ponderações de análise foram escaladas, de modo que elas não sejam incluídas ao tamanho da população. A estimação de variação SRS é usada no cálculo de estatísticas, como o efeito do design. A estimativa WR pode ser especificada apenas no estágio final de um design; o Assistente não permitirá a inclusão de outro estágio se a estimativa WR for selecionada.

WOR igual (amostragem de probabilidade igual sem substituição). A estimação WOR igual inclui a correção da população finita e assume que as unidades são amostradas com igual probabilidade. WOR igual pode ser especificado em qualquer estágio de um design.

WOR desigual (amostragem de probabilidade desigual sem substituição). Além de usar a correção da população finita, o WOR desigual conta com unidades de amostragem (geralmente clusters) selecionadas com probabilidade desigual. Este método de estimação está disponível apenas no primeiro estágio.

Assistente de preparação de análise: tamanho

Essa etapa é usada para especificar as probabilidades de inclusão ou tamanhos de população para o estágio atual. Os tamanhos podem ser fixos ou podem variar entre as camadas. Para efeitos de especificação de tamanhos, os clusters especificados nos estágios anteriores podem ser usados para definir as camadas. Observe que essa etapa é necessária apenas Equal WOR for escolhido como o Método de Estimação.

Unidades. É possível especificar tamanhos de população exatos ou as probabilidades com unidades que foram amostradas.

- **Valor.** Um valor único é aplicado a todas as camadas. Se **Tamanhos da população** for selecionado como a métrica de unidade, deve-se inserir um número inteiro não negativo. Se **Probabilidades de inclusão** for selecionado, deve-se digitar um valor entre 0 e 1, inclusivo.
- **Valores diferentes para as camadas.** Permite inserir valores de tamanho em uma base por estrato por meio da caixa de diálogo Definir tamanhos desiguais.
- **Valores de leitura da variável.** Permite selecionar uma variável numérica que contém valores de tamanho para as camadas.

Definir tamanhos desiguais

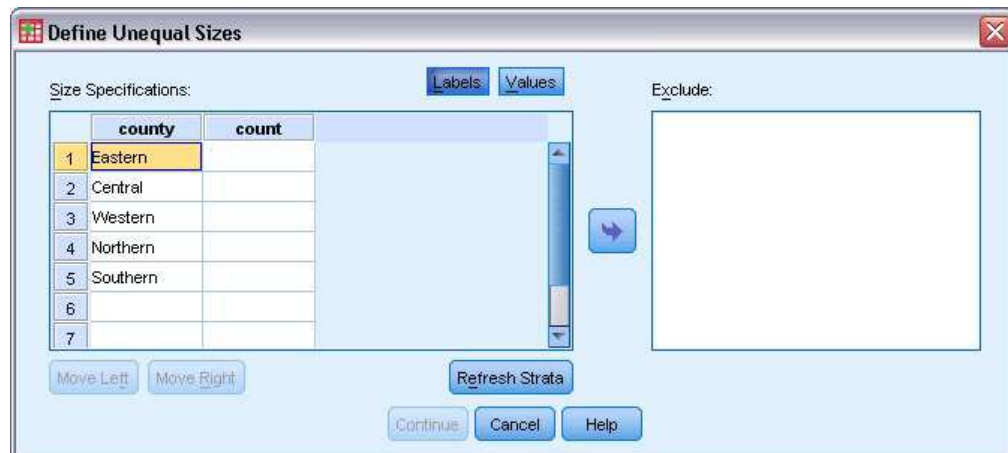


Figura 2. Caixa de diálogo Definir tamanhos desiguais

A caixa de diálogo Definir tamanhos desiguais permite inserir tamanhos em uma base por camada.

Grade de especificações de tamanho. A grade exibe as classificações cruzadas de até cinco variáveis de camada ou cluster — uma combinação de camada/cluster por linha. As variáveis de grade elegíveis incluem todas as variáveis de estratificação dos estágios atual e anterior e todas as variáveis de cluster de estágios anteriores. As variáveis podem ser reordenadas na grade ou movidas para a lista Exclusão. Insira tamanhos na coluna mais à direita. Clique em **Rótulos** ou **Valores** para alternar a exibição de rótulos de valor e de valores de dados para variáveis de estratificação e de cluster nas células de grade. As células que contêm valores não rotulados sempre mostram valores. Clique em **Atualizar camada** para preencher novamente a grade com cada combinação de valores de dados rotulados para variáveis na grade.

Exclusão. Para especificar tamanhos para um subconjunto de combinações de camada/cluster, mova uma ou mais variáveis para a lista Exclusão. Essas variáveis não são usadas para definir tamanhos de amostra.

Análise do assistente de preparação: resumo do plano

Esta é a última etapa dentro de cada estágio, fornecendo um resumo das especificações de design de análise através do estágio atual. A partir daqui, é possível prosseguir para o próximo estágio (criando-o se necessário) ou salvar as especificações de análise.

Se não for possível incluir outro estágio, isto provável se deve ao fato de:

- Nenhuma variável de cluster foi especificada na etapa de Designar Variáveis.
- A estimativa WR foi selecionada na etapa de método de estimação.
- Este é o terceiro estágio da análise e o Assistente suporta um máximo de três estágios.

Assistente de preparação de análise: conclusão

Esta é a etapa final. É possível salvar o arquivo de plano agora ou colar suas seleções em uma janela de sintaxe.

Ao fazer mudanças para estágios no arquivo de plano existente, é possível salvar o plano editado em um novo arquivo ou sobrescrever o arquivo existente. Ao incluir estágios sem fazer mudanças para estágios existente, o Assistente automaticamente sobrescreve o arquivo de plano existente. Se você deseja salvar o plano em um novo arquivo, escolha **Colar a sintaxe gerada pelo Assistente em uma janela de sintaxe** e altere o nome do arquivo nos comandos de sintaxe.

Modificando um plano de análise existente

1. Nos menus, escolha:
Analisar > Amostras complexas > Preparar para análise...
2. Selecione **Editar um arquivo de plano** e escolha o nome de um arquivo de plano para o qual você salvará o plano de análise.
3. Clique em **Avançar** para continuar pelo assistente.
4. Revise o plano de análise na etapa Resumo de Plano e, em seguida, clique em **Avançar**.
As etapas subsequentes são em grande parte as mesmas que para um novo design. Para obter mais informações, consulte a Ajuda para etapas individuais.
5. Navegue para a etapa Concluir e especifique um novo nome para o arquivo de plano editado ou escolha sobrescrever o arquivo de plano existente.

Opcionalmente, é possível remover estágios do plano.

Análise do assistente de preparação: resumo do plano

Esta etapa permite revisar o plano de análise e remover estágios do plano.

Remover estágios. É possível remover os estágios 2 e 3 de um design de vários estágios. Como um planejamento deve ter pelo menos um estágio, é possível editar, mas não remover o estágio 1 do design.

Capítulo 4. Plano de amostras complexo

Os procedimentos de análise de Amostras Complexas requerem especificações de análise a partir de uma análise ou de um arquivo de plano de amostra, a fim de fornecer resultados válidos.

Plano. Especifique o caminho de uma análise ou arquivo de plano de amostra.

Probabilidades conjuntas. Para usar a estimação WOR Desigual para clusters colhidos utilizando um método PPS WOR, é necessário especificar um arquivo separado ou um conjunto de dados aberto que contém as probabilidades conjuntas. Esse arquivo ou conjunto de dados é criado pelo Assistente de Amostragem durante a amostragem.

Capítulo 5. Frequências de amostras complexas

O procedimento Frequências de amostras complexas produz tabelas de frequência para variáveis selecionadas e exibe estatísticas univariadas. Opcionalmente, é possível solicitar estatísticas por subgrupos, definidos por uma ou mais variáveis categóricas.

Exemplo. Usando o procedimento Frequências de amostras complexas, é possível obter estatísticas tabulares univariadas para uso da vitamina entre os cidadãos dos EUA, com base nos resultados da National Health Interview Survey (NHIS) e com um plano de análise apropriado para esses dados de uso público.

Estatísticas. O procedimento produz estimativas de tamanhos da população da célula e porcentagens de tabela, além de erros padrão, intervalos de confiança, coeficientes de variação, efeitos de design, raízes quadradas de efeitos de design, valores acumulativos e contagens não ponderadas para cada estimativa. Além disso, as estatísticas qui-quadrado e as estatísticas de razão de verossimilhança são calculadas para o teste de proporções de células iguais.

Considerações de dados de Frequências de amostras complexas

Dados. As variáveis para as quais tabelas de frequências são produzidas devem ser categóricas. As variáveis de subpopulação podem ser de sequência de caracteres ou numéricas, mas devem ser categóricas.

Suposições. Os casos no arquivo de dados representam uma amostra a partir de um design complexo que deve ser analisado de acordo com as especificações no arquivo selecionado na Caixa de diálogo Plano de amostras complexas.

Obtendo Frequências de amostras complexas

1. Nos menus, escolha:
Analisar > Amostras complexas > Frequências...
2. Selecione um arquivo de plano. De forma opcional, selecione um arquivo de probabilidade conjunta customizado.
3. Clique em **Continuar**.
4. Na caixa de diálogo Frequências de amostras complexas, selecione pelo menos uma variável de frequência.

Opcionalmente, é possível especificar variáveis para definir subpopulações. As estatísticas são calculadas separadamente para cada subpopulação.

Estatísticas de frequências de amostras complexas

Células. Este grupo permite solicitar estimativas dos tamanhos da população da célula e porcentagens de tabela.

Estatísticas. Este grupo produz estatísticas associadas ao tamanho da população ou porcentagem da tabela.

- **Erro padrão.** O erro padrão da estimativa.
- **Intervalo de confiança.** Um intervalo de confiança para a estimativa, utilizando o nível especificado.
- **Coeficiente de variação.** A razão do erro padrão da estimativa para a estimativa.
- **Contagem não ponderada.** O número de unidades usadas para calcular a estimativa.

- **Efeito do design.** A razão da variância da estimativa com a variância obtida ao supor que a amostra seja uma amostra aleatória simples. Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.
- **Valores acumulativos.** A estimativa acumulativa ao longo de cada valor da variável.

Teste de proporções de célula iguais. Isso produz testes qui-quadrado e de razão de verossimilhança das hipóteses que as categorias de uma variável têm frequências iguais. Testes separados são executados para cada variável.

Valores omissos de amostras complexas

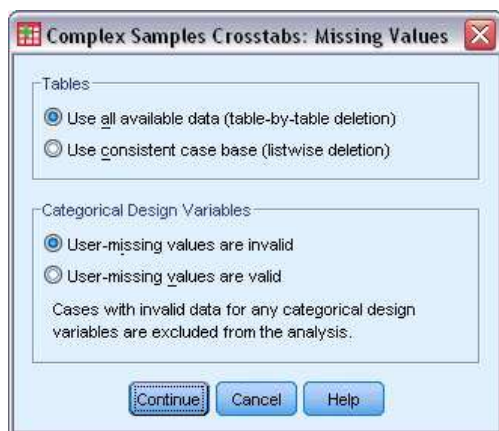


Figura 3. Caixa de diálogo Valores omissos

Tabelas. Este grupo determina quais casos são usados na análise.

- **Use todos os dados disponíveis.** Os valores omissos são determinados em uma base de tabela por tabela. Desta forma, os casos usados para calcular estatísticas podem variar entre tabelas de frequência e tabulação cruzada.
- **Usar base de caso consistente.** Os valores omissos são determinados em todas as variáveis. Desta forma, os casos usados para calcular estatísticas são consistentes entre tabelas.

Variáveis de design categórico. Este grupo determina se os valores omissos do usuário são válidos ou inválidos.

Opções de amostras complexas

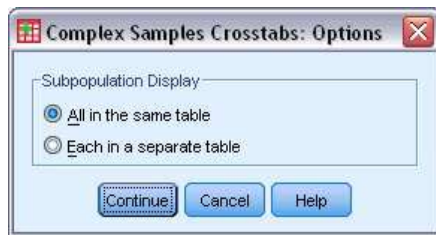


Figura 4. Caixa de diálogo Opções

Exibição de subpopulação. É possível escolher ter subpopulações exibidas na mesma tabela ou em tabelas separadas.

Capítulo 6. Descrições de amostras complexas

O procedimento Amostras complexas descritivas exibe estatísticas básicas univariadas para várias variáveis. Opcionalmente, é possível solicitar estatísticas por subgrupos, definidos por uma ou mais variáveis categóricas.

Exemplo. Usando o procedimento Amostras Complexas Descritivas, é possível obter estatísticas descritivas univariadas para os níveis de atividade de cidadãos dos EUA, com base nos resultados da National Health Entrevistar Survey (NHIS) e com um plano de análise apropriado para estes dados de uso público.

Estatísticas. O procedimento produz as médias e somas, mais testes t , erros padrão, intervalos de confiança, coeficientes de variação, contagens não ponderadas, tamanhos da população, efeitos de design e as raízes quadradas de efeitos de design para cada estimativa.

Considerações de dados de amostras complexas descritivas

Dados. As medidas devem ser variáveis de escala. As variáveis de subpopulação podem ser de sequência de caracteres ou numéricas, mas devem ser categóricas.

Suposições. Os casos no arquivo de dados representam uma amostra de um design complexo que deve ser analisado de acordo com as especificações no arquivo selecionado na caixa de diálogo Plano de amostras complexas.

Obtendo amostras complexas descritivas

1. Nos menus, escolha:
Analisar > Amostras complexas > Descritivas...
2. Selecione um arquivo de plano. De forma opcional, selecione um arquivo de probabilidade conjunta customizado.
3. Clique em **Continuar**.
4. Na caixa de diálogo Amostras complexas descritivas, selecione pelo menos uma variável de medida.

Opcionalmente, é possível especificar variáveis para definir subpopulações. As estatísticas são calculadas separadamente para cada subpopulação.

Estatísticas descritivas de amostras complexas

Resumos. Este grupo permite solicitar estimativas das médias das variáveis de medida. Além disso, é possível solicitar testes t das estimativas com relação a um valor especificado.

Estatísticas. Este grupo produz estatísticas associadas à média ou soma.

- **Erro padrão.** O erro padrão da estimativa.
- **Intervalo de confiança.** Um intervalo de confiança para a estimativa, utilizando o nível especificado.
- **Coeficiente de variação.** A razão do erro padrão da estimativa para a estimativa.
- **Contagem não ponderada.** O número de unidades usadas para calcular a estimativa.
- **Tamanho da população.** O número estimado de unidades na população.
- **Efeito do design.** A razão da variância da estimativa com a variância obtida ao supor que a amostra seja uma amostra aleatória simples. Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.

- **Raiz quadrada de efeito do design.** Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.

Valores omissos de amostras complexas descritivas

Estatísticas para variáveis de medida. Este grupo determina quais casos são usados na análise.

- **Use todos os dados disponíveis.** Os valores omissos são determinados em uma base de variável por variável, assim, os casos usados para calcular estatísticas podem variar entre as variáveis de medida.
- **Assegure a base de caso consistente.** Os valores omissos são determinados entre todas as variáveis, desta forma, os casos usados para calcular estatísticas são consistentes.

Variáveis de design categórico. Este grupo determina se os valores omissos do usuário são válidos ou inválidos.

Opções de amostras complexas

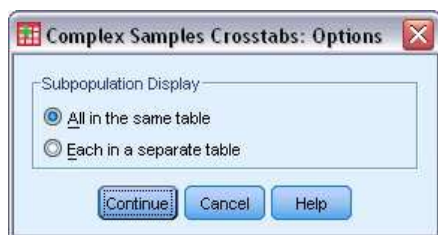


Figura 5. Caixa de diálogo Opções

Exibição de subpopulação. É possível escolher ter subpopulações exibidas na mesma tabela ou em tabelas separadas.

Capítulo 7. Tabulações cruzadas de amostras complexas

O procedimento Tabulações cruzadas de amostras complexas produz tabelas de tabulações cruzadas para pares de variáveis selecionadas e exibe estatísticas de dois fatores. Opcionalmente, é possível solicitar estatísticas por subgrupos, definidos por uma ou mais variáveis categóricas.

Exemplo. Usando o procedimento Tabulações cruzadas de amostras complexas, é possível obter estatísticas de classificação cruzada para frequência de fumar pelo uso de vitamina de cidadãos dos EUA, com base nos resultados da National Health Interview Survey (NHIS) e com um plano de análise apropriado para estes dados de uso público.

Estatísticas. O procedimento produz estimativas dos tamanhos da população da célula e linha, coluna, além de porcentagens de tabela, erros padrão, intervalos de confiança, coeficientes de variação, valores esperados, efeitos de design, raízes quadradas de efeitos de design, resíduos, resíduos ajustados e contagens ponderadas para cada estimativa. A razão de chances, risco relativo e a diferença de risco são calculados para tabelas 2 por 2. Além disso, as estatísticas de Pearson e de razão de verossimilhança são calculadas para o teste de independência das variáveis de linha e de coluna.

Considerações de dados de tabulações cruzadas de amostras complexas

Dados. As variáveis de linha e coluna devem ser categóricas. As variáveis de subpopulação podem ser de sequência de caracteres ou numéricas, mas devem ser categóricas.

Suposições. Os casos no arquivo de dados representam uma amostra de um design complexo que deve ser analisado de acordo com as especificações no arquivo selecionado na caixa de diálogo Plano de amostras complexas.

Obtendo tabulações cruzadas de amostras complexas

1. Nos menus, escolha:
Analisar > Amostras complexas > Tabulações cruzadas...
2. Selecione um arquivo de plano. De forma opcional, selecione um arquivo de probabilidade conjunta customizado.
3. Clique em **Continuar**.
4. Na caixa de diálogo Tabulações cruzadas de amostras complexas, selecione pelo menos uma variável de linha e uma variável da coluna.

Opcionalmente, é possível especificar variáveis para definir subpopulações. As estatísticas são calculadas separadamente para cada subpopulação.

Estatísticas de Tabulações cruzadas de amostras complexas

Células. Este grupo permite solicitar estimativas do tamanho da população da célula e as porcentagens de linha, coluna e tabela.

Estatísticas. Este grupo produz estatísticas associadas ao tamanho da população e às porcentagens de linha, coluna e tabela.

- **Erro padrão.** O erro padrão da estimativa.
- **Intervalo de confiança.** Um intervalo de confiança para a estimativa, utilizando o nível especificado.
- **Coefficiente de variação.** A razão do erro padrão da estimativa para a estimativa.

- **Valores esperados.** O valor esperado da estimativa, sob a hipótese de independência da linha e variável da coluna.
- **Contagem não ponderada.** O número de unidades usadas para calcular a estimativa.
- **Efeito do design.** A razão da variância da estimativa para a variância obtida ao supor que a amostra é uma amostra aleatória simples. Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.
- **Residuais.** O valor esperado é o número de casos que seria esperado na célula, se não houvesse nenhum relacionamento entre as duas variáveis. Um resíduo positivo indica que há mais casos na célula do que teria se as variáveis de linha e de coluna fossem independentes.
- **Resíduos ajustados.** O resíduo de uma célula (valor observado menos esperado) dividido por uma estimativa do seu erro padrão. O resíduo padronizado resultante é expresso em unidades de desvio padrão acima ou abaixo da média.

Resumos para tabelas 2 por 2. Este grupo produz estatísticas para tabelas, nas quais a variável de linha e coluna tem cada uma duas categorias. Cada uma é uma medida da intensidade da associação entre a presença de um fator e a ocorrência de um evento.

- **Razão de chances.** A razão de chances pode ser usada como uma estimativa de risco relativo quando a ocorrência do fator for rara.
- **Risco relativo.** A razão do risco de um evento na presença do fator com o risco do evento na ausência do fator.
- **Diferença de risco.** A diferença entre o risco de um evento na presença do fator e o risco do evento na ausência do fator.

Teste de independência de linhas e colunas. Isso produz testes de qui-quadrado e de razão de verossimilhança da hipótese de que uma variável de linha e de coluna são independentes. Testes separados são executados para cada par de variáveis.

Valores omissos de amostras complexas

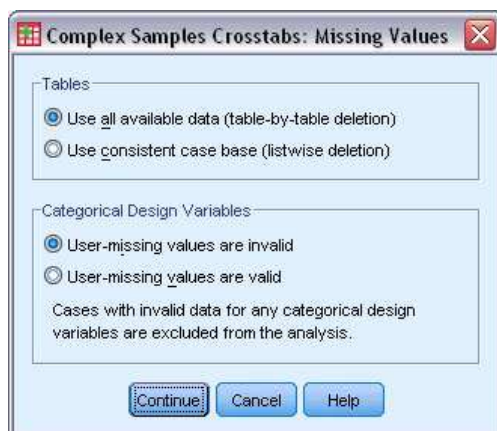


Figura 6. Caixa de diálogo Valores omissos

Tabelas. Este grupo determina quais casos são usados na análise.

- **Use todos os dados disponíveis.** Os valores omissos são determinados em uma base de tabela por tabela. Desta forma, os casos usados para calcular estatísticas podem variar entre tabelas de frequência e tabulação cruzada.

- **Usar base de caso consistente.** Os valores omissos são determinados em todas as variáveis. Desta forma, os casos usados para calcular estatísticas são consistentes entre tabelas.

Variáveis de design categórico. Este grupo determina se os valores omissos do usuário são válidos ou inválidos.

Opções de amostras complexas

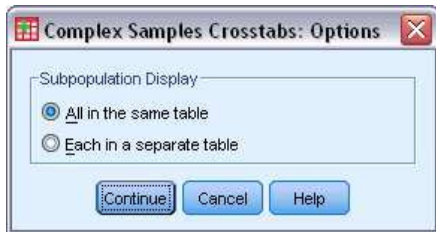


Figura 7. Caixa de diálogo Opções

Exibição de subpopulação. É possível escolher ter subpopulações exibidas na mesma tabela ou em tabelas separadas.

Capítulo 8. Razões de amostras complexas

O procedimento Razões de amostras complexas exibe estatísticas de resumo univariadas para razões de variáveis. Opcionalmente, é possível solicitar estatísticas por subgrupos, definidos por uma ou mais variáveis categóricas.

Exemplo. Usando o procedimento Razões de amostras complexas, é possível obter estatísticas descritivas para a razão do valor da propriedade atual para o último valor avaliado, com base nos resultados de uma pesquisa em todo o estado realizada de acordo com um design complexo e com um plano de análise apropriado para os dados.

Estatísticas. O procedimento produz estimativas de razão, testes t , erros padrão, intervalos de confiança, coeficientes de variação, contagem não ponderadas, tamanhos de população, efeitos do design e as raízes quadradas de efeitos do design.

Considerações de dados de proporção de amostras complexas

Dados. Numeradores e denominadores devem ser variáveis de escala de valor positivo. As variáveis de subpopulação podem ser de sequência de caracteres ou numéricas, mas devem ser categóricas.

Suposições. Os casos no arquivo de dados representam uma amostra de um design complexo que deve ser analisado de acordo com as especificações no arquivo selecionado na caixa de diálogo Plano de amostras complexas.

Obtendo razões de amostras complexas

1. Nos menus, escolha:
Analisar > Amostras complexas > Razões...
2. Selecione um arquivo de plano. De forma opcional, selecione um arquivo de probabilidade conjunta customizado.
3. Clique em **Continuar**.
4. Na caixa de diálogo Razões de amostras complexas, selecione pelo menos uma variável do numerador e uma variável do denominador.

Opcionalmente, é possível especificar variáveis para definir subgrupos para os quais as estatísticas são produzidas.

Estatísticas de razão de amostras complexas

Estatísticas. Este grupo produz estatísticas associadas à estimativa de razão.

- **Erro padrão.** O erro padrão da estimativa.
- **Intervalo de confiança.** Um intervalo de confiança para a estimativa, utilizando o nível especificado.
- **Coefficiente de variação.** A razão do erro padrão da estimativa para a estimativa.
- **Contagem não ponderada.** O número de unidades usadas para calcular a estimativa.
- **Tamanho da população.** O número estimado de unidades na população.
- **Efeito do design.** A razão da variância da estimativa com a variância obtida ao supor que a amostra seja uma amostra aleatória simples. Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.

teste t . É possível selecionar testes t das estimativas com relação a um valor especificado.

Valores omissos de razões de amostras complexas

Razões. Este grupo determina quais casos são usados na análise.

- **Use todos os dados disponíveis.** Os valores omissos são determinados em uma base razão por razão. Assim, os casos usados para calcular estatísticas podem variar entre pares de denominador-numerador.
- **Assegure a base de caso consistente.** Os valores omissos são determinados em todas as variáveis. Assim, os casos usados para calcular estatísticas são consistentes.

Variáveis de design categórico. Este grupo determina se os valores omissos do usuário são válidos ou inválidos.

Opções de amostras complexas

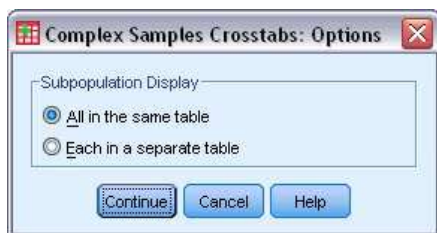


Figura 8. Caixa de diálogo Opções

Exibição de subpopulação. É possível escolher ter subpopulações exibidas na mesma tabela ou em tabelas separadas.

Capítulo 9. Modelo linear geral de amostras complexas

O procedimento Modelo linear geral de amostras complexas (CSGLM) executa análise de regressão linear, bem como análise de variância e covariância, para amostras colhidas por métodos de amostragem complexas. Opcionalmente, é possível solicitar análises para uma subpopulação.

Exemplo. Uma cadeia de supermercados entrevistou um conjunto de clientes sobre seus hábitos de compra, de acordo com um design complexo. Dados os resultados da pesquisa e quanto cada cliente gastou no mês anterior, a loja quer ver se a frequência com a qual os clientes compram está relacionada à quantia que ele gastam em um mês, fazendo um controle para o sexo do cliente e incorporando o plano de amostragem.

Estatísticas. O procedimento produz estimativas, erros padrão, intervalos de confiança, testes *t*, efeitos de design e raízes quadradas de efeitos de design para parâmetros do modelo, bem como as correlações e covariâncias entre estimativas paramétrica. As medidas de ajuste do modelo e as estatísticas descritivas para as variáveis dependentes e independentes também estão disponíveis. Além disso, é possível solicitar médias marginais estimadas para os níveis de fatores de modelo e interações entre fatores.

Considerações de dados do Modelo linear geral de amostras complexas

Dados. A variável dependente é quantitativa. Os fatores são categóricos. Covariáveis são variáveis quantitativas que estão relacionadas à variável dependente. As variáveis de subpopulação podem ser de sequência de caracteres ou numéricas, mas devem ser categóricas.

Suposições. Os casos no arquivo de dados representam uma amostra a partir de um design complexo que devem ser analisada de acordo com as especificações no arquivo selecionado na caixa de diálogo Plano de amostras complexas.

Obtendo um Modelo linear geral de amostras complexas

1. Nos menus, escolha:
Analisar > Amostras complexas > Modelo linear geral...
2. Selecione um arquivo de plano. De forma opcional, selecione um arquivo de probabilidade conjunta customizado.
3. Clique em **Continuar**.
4. Na caixa de diálogo Modelo linear geral de amostras complexas, selecione uma variável dependente.

Como opção, você pode:

- Selecione variáveis para fatores e covariáveis, conforme apropriado para seus dados.
- Especifique uma variável para definir uma subpopulação. A análise é executada apenas para a categoria selecionada da variável de subpopulação.

Modelo linear geral de amostras complexas

Especificar efeitos do modelo. Por padrão, o procedimento constrói um modelo dos principais efeitos usando os fatores e covariáveis especificados na caixa de diálogo principal. Como alternativa, é possível construir um modelo customizado que inclui efeitos de interação e termos aninhados.

Termos não aninhados

Para os fatores e covariáveis selecionados:

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

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

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

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

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

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

Termos aninhados

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

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

Limitações. Termos aninhados têm as restrições a seguir:

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

Intercepto. O intercepto geralmente é incluído no modelo. Se você conseguir presumir a passagem de dados por meio da origem, será possível excluir o intercepto. Mesmo se você incluir o intercepto no modelo, poderá escolher suprimir as estatísticas relacionados a ele.

Estatísticas de modelo linear geral de amostras complexas

Parâmetros de modelo. Este grupo permite controlar a exibição de estatísticas relacionadas aos parâmetros de modelo.

- **Estimativa.** Exibe estimativas dos coeficientes.
- **Erro padrão.** Exibe o erro padrão para cada estimativa de coeficiente.
- **Intervalo de confiança.** Exibe um intervalo de confiança para cada estimativa de coeficiente. O nível de confiança para o intervalo é configurado na caixa de diálogo Opções.
- **Teste T.** Exibe um teste t de cada estimativa de coeficiente. A hipótese nula para cada teste é que o valor do coeficiente é 0.
- **Covariâncias de estimativas paramétrica.** Exibe uma estimativa da matriz de covariâncias para os coeficientes do modelo.
- **Correlações de estimativas paramétrica.** Exibe uma estimativa da matriz de correlações para os coeficientes do modelo.
- **Efeito do design.** A razão da variância da estimativa com a variância obtida ao supor que a amostra seja uma amostra aleatória simples. Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.

- **Raiz quadrada de efeito do design.** Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.

Ajuste do modelo. Exibe estatísticas de R^2 e de erro quadrático médio raiz.

Médias de população de variável dependente e covariáveis. Exibe informações de sumarização sobre a variável dependente, covariáveis e fatores.

Informações do plano de amostra. Exibe informações de sumarização sobre a amostra, incluindo a contagem não ponderada e o tamanho da população.

Testes de hipótese de amostras complexas

Estatística do teste. Este grupo permite selecionar o tipo de estatística usada para testar hipóteses. É possível escolher entre F , F ajustado, qui-quadrado e qui-quadrado ajustado.

Amostragem de graus de liberdade. Este grupo fornece o controle sobre os graus de liberdade do plano de amostragem usados para calcular valores de p para todas as estatísticas de teste. Se baseado no plano de amostragem, o valor é a diferença entre o número de unidades de amostragem primária e o número de camadas no primeiro estágio de amostragem. Como alternativa, é possível configurar um grau de liberdade especificando um número inteiro positivo.

Ajuste para múltiplas comparações. Ao executar teste de hipóteses com diversos contrastes, o nível de significância global poderá ser ajustado a partir dos níveis de significância para os contrastes incluídos. Este grupo permite escolher o método de ajustamento.

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

Médias estimadas do modelo linear geral de amostras complexas

A caixa de diálogo Médias Estimadas permite exibir as médias marginais estimadas do modelo para os níveis de fatores e interações entre fatores especificados na caixa de subdiálogo Modelo. Também é possível solicitar que a média da população geral seja exibida.

Termo. Médias estimadas são calculadas para os fatores e interações entre fatores selecionados.

Contraste. O contraste determina como os testes de hipótese são configurados para comparar as médias estimadas.

- *Simples.* Compara a média de cada nível com a média de um nível especificado. Esse tipo de contraste é útil quando há um grupo de controle.
- *Desvio.* Compara a média de cada nível (exceto uma categoria de referência) com a média de todos os níveis (média geral). Os níveis do fator podem estar em qualquer ordem.
- *Diferença.* Compara a média de cada nível (exceto do primeiro) com a média dos níveis anteriores. Às vezes são chamados de contraste de Helmert reversos.

- *Helmert*. Compara a média de cada nível do fator (exceto do último) com a média dos níveis subsequentes.
- *Repetido*. Compara a média de cada nível (exceto do último) com a média dos níveis subsequentes.
- *Polinomial*. Compara o efeito linear, o efeito quadrático, o efeito cúbico, e assim por diante. O primeiro grau de liberdade contém o efeito linear em todas as categorias, o segundo grau de liberdade, o efeito quadrático, e assim por diante. Esses contrastes são frequentemente utilizados para estimar as tendências polinomiais.

Categoria de referência. Os contrastes simples e de desvio requerem uma categoria de referência ou nível do fator com relação aos outros aos quais são comparados.

Modelo linear geral de Amostras complexas - Salvar

Salvar variáveis. Esse grupo permite salvar os valores preditos e resíduos do modelo como novas variáveis no arquivo de trabalho.

Exportar modelo como dados do IBM SPSS Statistics. Grava um conjunto de dados no formato do IBM SPSS Statistics que contém a correlação paramétrica ou matriz de covariâncias com as estimativas paramétrica, erros padrão, valores de significância e graus de liberdade. A ordem das variáveis no arquivo matriz é conforme a seguir.

- **rowtype_.** Adota os valores (e rótulos de valor), COV (Covariâncias), CORR (Correlações), EST (estimativas paramétrica), SE (erros padrão), SIG (níveis de significância) e DF (graus de liberdade do design de amostragem). Há um caso separado com tipo de linha COV (ou CORR) para cada parâmetro de modelo, além de um caso separado para cada um dos outros tipos de linha.
- **varname_.** Adota os valores P1, P2,..., correspondendo a uma lista ordenada de todos os parâmetros de modelo para os tipos de linha COV ou CORR, com os rótulos de valor correspondentes às sequências paramétricas mostradas na tabela de estimativas paramétrica. As células estão em branco para outros tipos de linha.
- **P1, P2, ...** Essas variáveis correspondem a uma lista ordenada de todos os parâmetros de modelo, com os rótulos de variáveis correspondentes às sequências paramétricas mostradas na tabela de estimativas paramétrica, e adotam os valores de acordo com o tipo de linha. Para parâmetros redundantes, todas as covariâncias são definidas como zero; as correlações são configuradas como o valor omissivo do sistema; todas as estimativas paramétrica são definidas como zero; e todos os erros padrão, níveis de significância e graus de liberdade residuais são definidos como o valor omissivo do sistema.

Nota: este arquivo não é imediatamente utilizável para análises adicionais em outros procedimentos que leem um arquivo matriz, a menos que esses procedimentos aceitem todos os tipos de linhas exportados aqui.

Exportar modelo como XML. Salva as estimativas paramétrica e a matriz de covariância paramétrica, se selecionada, 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 modelo linear geral de amostras complexas

Valores omissos de usuário. Todas as variáveis de design, bem como a variável dependente e qualquer covariável, deve ter dados válidos. Casos com dados inválidos para qualquer uma dessas variáveis são excluídos da análise. Estes controles permitem decidir se valores omissos de usuário são tratados como válidos entre as camadas, cluster, subpopulação variáveis de fator.

Intervalo de Confiança. Este é o nível de intervalo de confiança para estimativas de coeficiente e médias marginais estimadas. Especifique um valor maior ou igual a 50 e menor que 100.

Recursos adicionais de comando CSGLM

O idioma da sintaxe de comando também permite:

- Especifique testes customizados de efeitos versus uma combinação linear ou um valor (usando o subcomando CUSTOM).
- Corrija as covariáveis em valores diferentes de suas médias ao calcular médias marginais estimadas (usando o subcomando EMMEANS).
- Especifique uma métrica para contrastes polinomiais (usando o subcomando EMMEANS).
- Especifique um valor de tolerância para a verificação de singularidade (usando o subcomando CRITERIA).
- Crie nomes especificados pelo usuário para as variáveis salvas (utilizando o subcomando SAVE).
- Produzir uma tabela de funções estimável geral (usando o subcomando PRINT).

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

Capítulo 10. Regressão logística de amostras complexas

O procedimento Regressão logística de amostras complexas executa análise de regressão logística em uma variável dependente binária ou multinomial para amostras exibidas por métodos de amostragem complexa. Opcionalmente, é possível solicitar análises para uma subpopulação.

Exemplo. Um responsável pelo empréstimo coletou registros passados de clientes que fizeram empréstimos em várias agências diferentes, de acordo com um design complexo. Ao incorporar o plano de amostra, o responsável deseja verificar se a probabilidade de um cliente dar calote está relacionada à idade, histórico de emprego e valor da dívida.

Estatísticas. O procedimento produz estimativas, estimativas exponenciadas, erros padrão, intervalos de confiança, testes t , efeitos de design e raízes quadradas de efeitos de design para parâmetros de modelo, bem como as correlações e covariâncias entre estimativas paramétrica. Pseudoestatísticas R^2 , tabelas de classificação e estatísticas descritivas para as variáveis dependentes e independentes também estão disponíveis.

Considerações de dados de regressão logística de amostras complexas

Dados. A variável dependente é categórica. Os fatores são categóricos. Covariáveis são variáveis quantitativas que estão relacionadas à variável dependente. As variáveis de subpopulação podem ser de sequência de caracteres ou numéricas, mas devem ser categóricas.

Suposições. Os casos no arquivo de dados representam uma amostra de um design complexo que deve ser analisada de acordo com as especificações no arquivo selecionado na caixa de diálogo Plano de amostras complexas.

Obtendo Regressão logística de amostras complexas

1. Nos menus, escolha:
Analisar > Amostras complexas > Regressão logística...
2. Selecione um arquivo de plano. De forma opcional, selecione um arquivo de probabilidade conjunta customizado.
3. Clique em **Continuar**.
4. Na caixa de diálogo Regressão logística de amostras complexas, selecione uma variável dependente.

Como opção, você pode:

- Selecione variáveis para fatores e covariáveis, conforme apropriado para seus dados.
- Especifique uma variável para definir uma subpopulação. A análise é executada apenas para a categoria selecionada da variável de subpopulação.

Categoria de referência de regressão logística de amostras complexas

Por padrão, o procedimento Regressão logística de amostras complexas torna a categoria de valor mais alto a categoria de referência. Essa caixa de diálogo permite especificar o valor mais alto, o valor mais baixo ou uma categoria customizada como a categoria de referência.

Modelo de regressão logística de Amostras Complexas

Especificar efeitos do modelo. Por padrão, o procedimento constrói um modelo dos principais efeitos usando os fatores e covariáveis especificados na caixa de diálogo principal. Como alternativa, é possível construir um modelo customizado que inclui efeitos de interação e termos aninhados.

Termos não aninhados

Para os fatores e covariáveis selecionados:

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

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

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

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

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

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

Termos aninhados

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

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

Limitações. Termos aninhados têm as restrições a seguir:

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

Intercepto. O intercepto geralmente é incluído no modelo. Se você conseguir presumir a passagem de dados por meio da origem, será possível excluir o intercepto. Mesmo se você incluir o intercepto no modelo, poderá escolher suprimir as estatísticas relacionados a ele.

Estatísticas de Regressão logística de amostras complexas

Ajuste do modelo. Controla a exibição de estatísticas que medem o desempenho geral do modelo.

- **Pseudo R-quadrado.** A estatística R^2 de regressão linear não possui uma contraparte exata entre modelos de regressão logística. Em vez disso, há várias medidas que tentam imitar as propriedades da estatística R^2 .
- **Tabela de classificação.** Exibe as classificações cruzadas tabuladas da categoria observada pela categoria predita do modelo na variável dependente.

Parâmetros. Este grupo permite controlar a exibição de estatísticas relacionadas aos parâmetros de modelo.

- **Estimativa.** Exibe estimativas dos coeficientes.
- **Estimativa exponenciada.** Exibe a base do logaritmo natural elevado à potência das estimativas dos coeficientes. Embora a estimativa tenha ótimas propriedades para teste estatístico, a estimativa exponenciada, ou $\exp(B)$, é mais fácil de interpretar.
- **Erro padrão.** Exibe o erro padrão para cada estimativa de coeficiente.
- **Intervalo de confiança.** Exibe um intervalo de confiança para cada estimativa de coeficiente. O nível de confiança para o intervalo é configurado na caixa de diálogo Opções.
- **Teste T.** Exibe um teste t de cada estimativa de coeficiente. A hipótese nula para cada teste é que o valor do coeficiente é 0.
- **Covariâncias de estimativas paramétrica.** Exibe uma estimativa da matriz de covariâncias para os coeficientes do modelo.
- **Correlações de estimativas paramétrica.** Exibe uma estimativa da matriz de correlações para os coeficientes do modelo.
- **Efeito do design.** A razão da variância da estimativa com a variância obtida ao supor que a amostra seja uma amostra aleatória simples. Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.

Estatísticas de sumarização para variáveis de modelo. Exibe informações de sumarização sobre a variável dependente, covariáveis e fatores.

Informações do plano de amostra. Exibe informações de sumarização sobre a amostra, incluindo a contagem não ponderada e o tamanho da população.

Testes de hipótese de amostras complexas

Estatística do teste. Este grupo permite selecionar o tipo de estatística usada para testar hipóteses. É possível escolher entre F , F ajustado, qui-quadrado e qui-quadrado ajustado.

Amostragem de graus de liberdade. Este grupo fornece o controle sobre os graus de liberdade do plano de amostragem usados calcular valores de p para todas as estatísticas de teste. Se baseado no plano de amostragem, o valor é a diferença entre o número de unidades de amostragem primária e o número de camadas no primeiro estágio de amostragem. Como alternativa, é possível configurar um grau de liberdade especificando um número inteiro positivo.

Ajuste para múltiplas comparações. Ao executar teste de hipóteses com diversos contrastes, o nível de significância global poderá ser ajustado a partir dos níveis de significância para os contrastes incluídos. Este grupo permite escolher o método de ajustamento.

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

Razões de chances de regressão logística de amostras complexas

A caixa de diálogo Razões de chances permite exibir razões de chances estimadas pelo modelo para fatores e covariáveis especificados. Um conjunto separado de razões de chances é calculado para cada categoria da variável dependente, exceto a categoria de referência.

Fatores. Para cada fator selecionado, exibe a razão das chances em cada categoria do fator para as chances na categoria de referência especificada.

Covariáveis. Para cada covariável selecionada, exibe a razão das chances no valor médio da covariável mais as unidades especificadas de mudança para as chances na média.

Ao calcular razões de chances para um fator ou covariável, o procedimento corrige todos os outros fatores em seus níveis mais altos e todas as outras covariáveis em suas médias. Se um fator ou covariável interagir com outros preditores no modelo, as razões de chances dependerão não apenas da mudança na variável especificada, mas também dos valores das variáveis com as quais ele interage. Se uma covariável especificada interagir com ela mesma no modelo (por exemplo, $age*age$), as razões de chances dependerão da mudança na covariável e do valor da covariável.

Regressão logística de amostras complexas - Salvar

Salvar variáveis. Esse grupo permite salvar a categoria predita do modelo e as probabilidades previstas como novas variáveis no conjunto de dados ativo.

Exportar modelo como dados do IBM SPSS Statistics. Grava um conjunto de dados no formato do IBM SPSS Statistics que contém a correlação paramétrica ou matriz de covariâncias com as estimativas paramétrica, erros padrão, valores de significância e graus de liberdade. A ordem das variáveis no arquivo matriz é conforme a seguir.

- **rowtype_.** Adota os valores (e rótulos de valor), COV (Covariâncias), CORR (Correlações), EST (estimativas paramétrica), SE (erros padrão), SIG (níveis de significância) e DF (graus de liberdade do design de amostragem). Há um caso separado com tipo de linha COV (ou CORR) para cada parâmetro de modelo, além de um caso separado para cada um dos outros tipos de linha.
- **varname_.** Adota os valores P1, P2,..., correspondendo a uma lista ordenada de todos os parâmetros de modelo para os tipos de linha COV ou CORR, com os rótulos de valor correspondentes às sequências paramétricas mostradas na tabela de estimativas paramétrica. As células estão em branco para outros tipos de linha.
- **P1, P2, ...** Essas variáveis correspondem a uma lista ordenada de todos os parâmetros de modelo, com os rótulos de variáveis correspondentes às sequências paramétricas mostradas na tabela de estimativas paramétrica, e adotam os valores de acordo com o tipo de linha. Para parâmetros redundantes, todas as covariâncias são definidas como zero; as correlações são configuradas como o valor omissivo do sistema; todas as estimativas paramétrica são definidas como zero; e todos os erros padrão, níveis de significância e graus de liberdade residuais são definidos como o valor omissivo do sistema.

Nota: este arquivo não é imediatamente utilizável para análises adicionais em outros procedimentos que leem um arquivo matriz, a menos que esses procedimentos aceitem todos os tipos de linhas exportados aqui.

Exportar modelo como XML. Salva as estimativas paramétrica e a matriz de covariância paramétrica, se selecionada, 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 de amostras complexas

Estimação. Esse grupo fornece controle de vários critérios na estimação de modelo.

- **Máximo de iterações.** O número máximo de iterações que o algoritmo executará. Especifique um número inteiro não negativo.
- **Divisão máxima do passo pela metade.** Em cada iteração, o tamanho do passo é reduzido por um fator de 0,5 até que o log da verossimilhança aumente ou a divisão máxima da etapa pela metade seja atingida. Especifique um número inteiro positivo.
- **Limitar iterações com base na mudança nas estimativas paramétrica.** Quando selecionada, o algoritmo para após uma iteração na qual a mudança absoluta ou relativa nas estimativas paramétrica é menor que o valor especificado, que deve ser não negativo.
- **Limitar iterações com base na mudança no log da verossimilhança.** Quando selecionada, o algoritmo para após uma iteração na qual a mudança absoluta ou relativa na função de log da verossimilhança é menor que o valor especificado, que deve ser não negativo.
- **Verificar separação completa de pontos de dados.** Quando selecionada, o algoritmo executa testes para assegurar que as estimativas do parâmetro tenham valores exclusivos. A separação ocorre quando o procedimento pode produzir um modelo que classifica corretamente cada caso.
- **Exibir histórico de iteração.** Exibe estimativas paramétrica e estatísticas a cada n iterações, começando com a 0th iteração (as estimativas iniciais). Se optar por imprimir o histórico de iteração, a última iteração será sempre impressa, independentemente do valor de n .

Valores omissos de usuário. Todas as variáveis de design, bem como a variável dependente e quaisquer covariáveis, devem ter dados válidos. Casos com dados inválidos para qualquer uma dessas variáveis são excluídos da análise. Esses controles permitem decidir se os valores omissos de usuário serão tratados como válidos entre camada, cluster, subpopulação e variáveis de fator.

Intervalo de Confiança. Esse é o nível de intervalo de confiança para estimativas de coeficiente, estimativas de coeficiente exponenciadas e razões de chances. Especifique um valor maior ou igual a 50 e menor que 100.

Recursos adicionais do comando CSLOGISTIC

O idioma da sintaxe de comando também permite:

- Especifique testes customizados de efeitos versus uma combinação linear ou um valor (usando o subcomando CUSTOM).
- Corrija os valores de outras variáveis de modelo ao calcular razões de chances para fatores e covariáveis (utilizando o subcomando ODDS RATIOS).
- Especifique um valor de tolerância para a verificação de singularidade (usando o subcomando CRITERIA).
- Crie nomes especificados pelo usuário para as variáveis salvas (utilizando o subcomando SAVE).
- Produzir uma tabela de funções estimável geral (usando o subcomando PRINT).

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

Capítulo 11. Regressão ordinal de amostras complexas

O procedimento Regressão ordinal de amostras complexas executa análise de regressão em uma variável dependente binária ou ordinal para amostras exibidas por métodos de amostragem complexa. Opcionalmente, é possível solicitar análises para uma subpopulação.

Exemplo. Os representantes que consideram um projeto de lei antes da legislatura estão interessados em se haverá suporte público para o projeto de lei e como o suporte para esse projeto está relacionado aos demográficos votantes. Os institutos de pesquisa projetam e conduzem entrevistas de acordo com um plano de amostragem complexa. Usando a Regressão ordinal de amostras complexas, é possível ajustar um modelo para o nível de suporte para o projeto de lei com base em demográficos votantes.

Considerações de dados de regressão ordinal de amostras complexas

Dados. A variável dependente é ordinal. Os fatores são categóricos. Covariáveis são variáveis quantitativas que estão relacionadas à variável dependente. As variáveis de subpopulação podem ser de sequência de caracteres ou numéricas, mas devem ser categóricas.

Suposições. Os casos no arquivo de dados representam uma amostra de um design complexo que deve ser analisado de acordo com as especificações no arquivo selecionado na caixa de diálogo Plano de amostras complexas.

Obtendo a regressão ordinal de amostras complexas

1. Nos menus, escolha:
Analisar > Amostras complexas > Regressão ordinal...
2. Selecione um arquivo de plano. De forma opcional, selecione um arquivo de probabilidade conjunta customizado.
3. Clique em **Continuar**.
4. Na caixa de diálogo Regressão ordinal de amostras complexas, selecione uma variável dependente.

Como opção, você pode:

- Selecione variáveis para fatores e covariáveis, conforme apropriado para seus dados.
- Especifique uma variável para definir uma subpopulação. A análise é executada somente para a categoria selecionada da variável de subpopulação, embora as variâncias ainda sejam estimadas corretamente com base no conjunto de dados inteiro.
- Selecione uma função de ligação.

Função de ligação. A função de ligação é uma transformação da probabilidades acumulativas que permite estimação do modelo. As cinco funções de link a seguir estão disponíveis.

- **Logit.** $f(x)=\log(x / (1-x))$. Geralmente usada para categorias uniformemente distribuídas.
- **Log-log complementar.** $f(x)=\log(-\log(1-x))$. Geralmente usada quando categorias superiores são mais prováveis.
- **Log-log negativo.** $f(x)=-\log(-\log(x))$. Geralmente usada quando categorias inferiores são mais prováveis.
- **Probit.** $f(x)=\Phi^{-1}(x)$. Geralmente usada quando a variável latente é normalmente distribuída.
- **Cauchit (Cauchy inverso).** $f(x) = \tan(\pi (x - 0.5))$. Geralmente usada quando a variável latente possui muitos valores extremos.

Probabilidades de resposta de Regressão ordinal de amostras complexas

A caixa de diálogo Probabilidades de resposta permite especificar se a probabilidade acumulativa de uma resposta (ou seja, a probabilidade de pertencer a e incluindo uma determinada categoria da variável dependente) aumenta com o aumento ou diminuição de valores da variável dependente.

Modelo de regressão ordinal de amostras complexas

Especificar efeitos do modelo. Por padrão, o procedimento constrói um modelo dos principais efeitos usando os fatores e covariáveis especificados na caixa de diálogo principal. Como alternativa, é possível construir um modelo customizado que inclui efeitos de interação e termos aninhados.

Termos não aninhados

Para os fatores e covariáveis selecionados:

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

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

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

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

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

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

Termos aninhados

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

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

Limitações. Termos aninhados têm as restrições a seguir:

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

Estatísticas de Regressão ordinal de Amostras complexas

Ajuste do modelo. Controla a exibição de estatísticas que medem o desempenho geral do modelo.

- **Pseudo R-quadrado.** A estatística R^2 de regressão linear não possui uma contraparte exata entre modelos de regressão linear. Em vez disso, há várias medidas que tentam imitar as propriedades da estatística R^2 .

- **Tabela de classificação.** Exibe as classificações cruzadas tabuladas da categoria observada pela categoria predita do modelo na variável dependente.

Parâmetros. Este grupo permite controlar a exibição de estatísticas relacionadas aos parâmetros de modelo.

- **Estimativa.** Exibe estimativas dos coeficientes.
- **Estimativa exponenciada.** Exibe a base do logaritmo natural elevado à potência das estimativas dos coeficientes. Embora a estimativa tenha ótimas propriedades para teste estatístico, a estimativa exponenciada, ou $\exp(B)$, é mais fácil de interpretar.
- **Erro padrão.** Exibe o erro padrão para cada estimativa de coeficiente.
- **Intervalo de confiança.** Exibe um intervalo de confiança para cada estimativa de coeficiente. O nível de confiança para o intervalo é configurado na caixa de diálogo Opções.
- **Teste T.** Exibe um teste t de cada estimativa de coeficiente. A hipótese nula para cada teste é que o valor do coeficiente é 0.
- **Covariâncias de estimativas paramétrica.** Exibe uma estimativa da matriz de covariâncias para os coeficientes do modelo.
- **Correlações de estimativas paramétrica.** Exibe uma estimativa da matriz de correlações para os coeficientes do modelo.
- **Efeito do design.** A razão da variância da estimativa com a variância obtida ao supor que a amostra seja uma amostra aleatória simples. Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Esta é uma medida, expressa em unidades comparáveis às do erro padrão, do efeito de especificar um design complexo, em que os valores mais distantes de 1 indicam maiores efeitos.

Linhas paralelas. Esse grupo permite solicitar estatísticas associadas a um modelo com linhas não paralelas em que uma linha de regressão separada é ajustada para cada categoria de resposta (exceto a última).

- **Teste de Wald.** Produz um teste da hipótese nula de que os parâmetros de regressão são iguais para todas as respostas acumulativas. O modelo com linhas não paralelas é estimado e o teste de Wald paramétricas iguais é aplicado.
- **Estimativas paramétrica.** Exibe estimativas dos coeficientes e os erros padrão para o modelo com linhas não paralelas.
- **Covariâncias de estimativas paramétrica.** Exibe uma estimativa da matriz de covariâncias para os coeficientes do modelo com linhas não paralelas.

Estatísticas de sumarização para variáveis de modelo. Exibe informações de sumarização sobre a variável dependente, covariáveis e fatores.

Informações do plano de amostra. Exibe informações de sumarização sobre a amostra, incluindo a contagem não ponderada e o tamanho da população.

Testes de hipótese de amostras complexas

Estatística do teste. Este grupo permite selecionar o tipo de estatística usada para testar hipóteses. É possível escolher entre F , F ajustado, qui-quadrado e qui-quadrado ajustado.

Amostragem de graus de liberdade. Este grupo fornece o controle sobre os graus de liberdade do plano de amostragem usados para calcular valores de p para todas as estatísticas de teste. Se baseado no plano de amostragem, o valor é a diferença entre o número de unidades de amostragem primária e o número de camadas no primeiro estágio de amostragem. Como alternativa, é possível configurar um grau de liberdade especificando um número inteiro positivo.

Ajuste para múltiplas comparações. Ao executar teste de hipóteses com diversos contrastes, o nível de significância global poderá ser ajustado a partir dos níveis de significância para os contrastes incluídos. Este grupo permite escolher o método de ajustamento.

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

Razões de chances de regressão ordinal de amostras complexas

A caixa de diálogo Razões de chances permite exibir razões de chances cumulativas estimadas pelo modelo para fatores e covariáveis especificados. Esse recurso está disponível somente para modelos que usam a função de ligação Logit. Uma única razão de chances acumulativas é calculada para todas as categorias da variável dependente, exceto a última; o modelo de chances proporcionais requer que todas sejam iguais.

Fatores. Para cada fator selecionado, exibe a razão das chances acumulativas em cada categoria do fator para as chances na categoria de referência especificada.

Covariáveis. Para cada covariável selecionada, exibe a razão das chances acumulativas no valor médio da covariável, além das unidades de mudança especificadas para as chances na média.

Ao calcular razões de chances para um fator ou covariável, o procedimento corrige todos os outros fatores em seus níveis mais altos e todas as outras covariáveis em suas médias. Se um fator ou covariável interagir com outros preditores no modelo, as razões de chances dependerão não apenas da mudança na variável especificada, mas também dos valores das variáveis com as quais ele interage. Se uma covariável especificada interagir com ela mesma no modelo (por exemplo, $age*age$), as razões de chances dependerão da mudança na covariável e do valor da covariável.

Regressão ordinal de amostras complexas - Salvar

Salvar variáveis. Esse grupo permite salvar a categoria predita do modelo, a probabilidade de categoria predita, a probabilidade de categoria observada, probabilidades acumulativas e as probabilidades preditas como novas variáveis no conjunto de dados ativo.

Exportar modelo como dados do IBM SPSS Statistics. Grava um conjunto de dados no formato do IBM SPSS Statistics que contém a correlação paramétrica ou matriz de covariâncias com as estimativas paramétrica, erros padrão, valores de significância e graus de liberdade. A ordem das variáveis no arquivo matriz é conforme a seguir.

- **rowtype_.** Adota os valores (e rótulos de valor), COV (Covariâncias), CORR (Correlações), EST (estimativas paramétrica), SE (erros padrão), SIG (níveis de significância) e DF (graus de liberdade do design de amostragem). Há um caso separado com tipo de linha COV (ou CORR) para cada parâmetro de modelo, além de um caso separado para cada um dos outros tipos de linha.
- **varname_.** Adota os valores P1, P2,..., correspondendo a uma lista ordenada de todos os parâmetros de modelo para os tipos de linha COV ou CORR, com os rótulos de valor correspondentes às sequências paramétricas mostradas na tabela de estimativas paramétrica. As células estão em branco para outros tipos de linha.

- **P1, P2, ...** Essas variáveis correspondem a uma lista ordenada de todos os parâmetros de modelo, com os rótulos de variáveis correspondentes às sequências paramétricas mostradas na tabela de estimativas paramétrica, e adotam os valores de acordo com o tipo de linha. Para parâmetros redundantes, todas as covariâncias são definidas como zero; as correlações são configuradas como o valor omisso do sistema; todas as estimativas paramétrica são definidas como zero; e todos os erros padrão, níveis de significância e graus de liberdade residuais são definidos como o valor omisso do sistema.

Nota: este arquivo não é imediatamente utilizável para análises adicionais em outros procedimentos que leem um arquivo matriz, a menos que esses procedimentos aceitem todos os tipos de linhas exportados aqui.

Exportar modelo como XML. Salva as estimativas paramétrica e a matriz de covariância paramétrica, se selecionada, 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 ordinal de amostras complexas

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

Estimação. Esse grupo fornece controle de vários critérios na estimação de modelo.

- **Máximo de iterações.** O número máximo de iterações que o algoritmo executará. Especifique um número inteiro não negativo.
- **Divisão máxima do passo pela metade.** Em cada iteração, o tamanho do passo é reduzido por um fator de 0,5 até que o log da verossimilhança aumente ou a divisão máxima da etapa pela metade seja atingida. Especifique um número inteiro positivo.
- **Limitar iterações com base na mudança nas estimativas paramétrica.** Quando selecionada, o algoritmo para após uma iteração na qual a mudança absoluta ou relativa nas estimativas paramétrica é menor que o valor especificado, que deve ser não negativo.
- **Limitar iterações com base na mudança no log da verossimilhança.** Quando selecionada, o algoritmo para após uma iteração na qual a mudança absoluta ou relativa na função de log da verossimilhança é menor que o valor especificado, que deve ser não negativo.
- **Verificar separação completa de pontos de dados.** Quando selecionada, o algoritmo executa testes para assegurar que as estimativas do parâmetro tenham valores exclusivos. A separação ocorre quando o procedimento pode produzir um modelo que classifica corretamente cada caso.
- **Exibir histórico de iteração.** Exibe estimativas paramétrica e estatísticas a cada n iterações, começando com a iteração 0 (as estimativas iniciais). Se optar por imprimir o histórico de iteração, a última iteração será sempre impressa, independentemente do valor de n .

Valores omissos de usuário. As variáveis de design de escala, bem como a variável dependente e quaisquer covariáveis, devem ter dados válidos. Casos com dados inválidos para qualquer uma dessas variáveis são excluídos da análise. Estes controles permitem decidir se valores omissos de usuário são tratados como válidos entre as camadas, cluster, subpopulação variáveis de fator.

Intervalo de Confiança. Esse é o nível de intervalo de confiança para estimativas de coeficiente, estimativas de coeficiente exponenciadas e razões de chances. Especifique um valor maior ou igual a 50 e menor que 100.

Recursos adicionais de comando CSORDINAL

O idioma da sintaxe de comando também permite:

- Especifique testes customizados de efeitos versus uma combinação linear ou um valor (usando o subcomando CUSTOM).
- Corrija os valores de outras variáveis de modelo em valores diferentes de suas médias ao calcular razões de chances cumulativas para fatores e covariáveis (usando o subcomando ODDSRATIOS).
- Use valores não rotulados como categorias de referência customizada para fatores quando razões de chances são solicitadas (usando o subcomando ODDSRATIOS).
- Especifique um valor de tolerância para a verificação de singularidade (usando o subcomando CRITERIA).
- Produzir uma tabela de funções estimável geral (usando o subcomando PRINT).
- Salvar mais de 25 variáveis de probabilidade (usando o subcomando SAVE).

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

Capítulo 12. Regressão de Cox de amostras complexas

O procedimento Regressão de Cox de amostras complexas executa análise de sobrevivência para amostras colhidas por métodos de amostragem complexas. Opcionalmente, é possível solicitar análises para uma subpopulação.

Exemplos. Uma agência de cumprimento da lei do governo está preocupada com taxas de reincidência na sua área de jurisdição. Uma das medidas de reincidência é o tempo até segunda prisão para os infratores. A agência gostaria de modelar o tempo para nova prisão usando a Regressão de Cox, mas está preocupada com a suposição de riscos proporcionais ser inválida entre as categorias de idade.

Os pesquisadores médicos estão investigando tempos de sobrevivência para pacientes que saem de um programa de reabilitação de acidente vascular cerebral pós-isquêmico. Existe o potencial para múltiplos casos por assunto, visto que os históricos dos pacientes mudam à medida que a ocorrência de eventos de não morte significativos são observadas e os tempos destes eventos registrados. A amostra também é truncada da esquerda no sentido de que os tempos de sobrevivência observados são "aumentados" pela duração da reabilitação, porque enquanto o início do risco é iniciado no momento do acidente vascular cerebral isquêmico, apenas os pacientes que sobrevivem após o programa de reabilitação estão na amostra.

Considerações dos dados de regressão de Cox de amostras complexas

Tempo de sobrevivência. O procedimento aplica a regressão de Cox para a análise dos tempos de sobrevivência—ou seja, o período de tempo antes da ocorrência de um evento. Existem duas maneiras de especificar o tempo de sobrevivência, dependendo do horário de início do intervalo:

- **Tempo=0.** Geralmente, você terá as informações completas sobre o início do intervalo para cada sujeito e terá simplesmente uma variável que contém horários de encerramentos (ou a criação de uma variável única com horários de encerramentos a partir das variáveis Data & Hora; veja abaixo).
- **Varia por sujeito.** Isso é apropriado quando você tem **truncamento da esquerda**, também chamado de **entrada atrasada**; por exemplo, se você estiver analisando os tempos de sobrevivência para doentes que estão saindo de um programa de reabilitação pós-AVC, é possível considerar que o início do risco é iniciado no momento do AVC. No entanto, se sua amostra incluir somente pacientes que sobreviveram ao programa de reabilitação, então sua amostra é truncada da esquerda no sentido de que os tempos de sobrevivência observados são "aumentados" pela duração da reabilitação. É possível considerar isso especificando a hora em que eles saírem da reabilitação como o tempo de entrada no estudo.

Variáveis de Data & Hora. As variáveis de Data & Hora não podem ser usadas para definir diretamente o início e o final do intervalo; se você tiver variáveis de Data & Hora, deve usá-las para criar variáveis que contém os tempos de sobrevivência. Se não houver truncamento da esquerda, simplesmente crie uma variável que contém os horários de encerramento com base na diferença entre a data de entrada no estudo e a data da observação. Se houver truncamento da esquerda, crie uma variável que contém os horários de início, com base na diferença entre a data do início no estudo e a data de entrada, além de uma variável que contém horários de encerramento, com base na diferença entre a data do início do estudo e a data de observação.

Status do evento. Uma variável que registra se o sujeito que experimentou o evento de interesse no intervalo é necessária. Sujeitos para os quais o evento não ocorreu estão censurados à direita.

Identificador de sujeito. É possível incorporar facilmente preditores dependentes de tempo, constantes em trechos, dividindo as observações para um único sujeito em vários casos. Por exemplo, se você estiver analisando tempos de sobrevivência para pacientes pós-AVC, variáveis que representam seus históricos

médicos devem ser úteis como preditores. Com o passar do tempo, eles podem passar por eventos médicos graves que alteram o seu histórico médico. A tabela a seguir mostra como estruturar esse conjunto de dados: *ID do Paciente* é o identificador do sujeito, *Horário de encerramento* define os intervalos observados, *Status* registra eventos médicos graves e *Histórico de ataque cardíaco anterior* e *Histórico de hemorragia anterior* são preditores dependentes de tempo constantes em trechos.

Tabela 1. Estrutura de dados para incorporar preditores dependentes de tempo constantes em trechos.

<i>ID do paciente</i>	<i>Horário de encerramento</i>	<i>Status</i>	<i>Histórico de ataque cardíaco anterior</i>	<i>Histórico de hemorragia anterior</i>
1	5	Ataque cardíaco	Não	Não
1	7	Hemorragia	Sim	Não
1	8	Morto	Sim	Sim
2	24	Morto	Não	Não
3	8	Ataque cardíaco	Não	Não
3	15	Morto	Sim	Não

Suposições. Os casos no arquivo de dados representam uma amostra a partir de um design complexo que deve ser analisado de acordo com as especificações no arquivo selecionado na Caixa de diálogo Plano de amostras complexas.

Geralmente, modelos de regressão de Cox supõem riscos proporcionais –ou seja, a razão dos riscos de um caso para outro não deve variar ao longo do tempo. Se esta presunção não se mantiver, pode ser necessário incluir preditores dependentes de tempo no modelo.

Análise de Kaplan-Meier. Se você não selecionar quaisquer preditores (ou não inserir quaisquer preditores selecionados no modelo) e escolher o método de limite do produto para calcular a curva de sobrevivência de linha de base na guia Opções, o procedimento executa um tipo Kaplan-Meier de análise de sobrevivência.

Para obter regressão de Cox de amostras complexas

1. Nos menus, escolha:
Analisar > Amostras complexas > Regressão de Cox...
2. Selecione um arquivo de plano. De forma opcional, selecione um arquivo de probabilidade conjunta customizado.
3. Clique em **Continuar**.
4. Especifique o tempo de sobrevivência, selecionando os horários de entrada e saída do estudo.
5. Selecione uma variável de status do evento.
6. Clique em Definir evento e defina pelo menos um valor do evento.

Opcionalmente, é possível selecionar um identificador de sujeito.

Definir evento

Especifique os valores que indicam que um evento terminal ocorreu.

- **Valor(es) individual(is).** Especifique um ou mais valores inserindo-os na grade ou selecionando-os a partir de uma lista de valores com os rótulos de valores definidos.
- **Intervalo de valores.** Especifique um intervalo de valores inserindo os valores mínimo e máximo ou selecionando valores a partir de uma lista com os rótulos de valores definidos.

Preditores

A guia Preditores permite especificar os fatores e covariáveis usados para construir efeitos do modelo.

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

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

Preditores dependentes de tempo. Há determinadas situações em que a suposição de riscos proporcionais não é mantida. Ou seja, as razões de risco ao longo do tempo de mudança; os valores de um (ou mais) de seus preditores são diferentes em diferentes pontos do tempo. Em tais casos, é necessário especificar preditores dependentes de tempo. Consulte o tópico “Definir preditor dependente de tempo” para obter mais informações. Os preditores dependentes de tempo podem ser selecionados como fatores ou covariáveis.

Definir preditor dependente de tempo

A caixa de diálogo Definir preditor dependente de tempo permite criar um preditor que é dependente da variável tempo integrada, $T_$. É possível usar esta variável para definir covariáveis dependentes de tempo de duas maneiras gerais:

- Se você deseja estimar um modelo de regressão de Cox estendido que permita riscos não proporcionais, será possível fazer isso definindo seu preditor dependentes de tempo como uma função da variável de tempo $T_$ e a covariável em questão. Um exemplo comum seria o produto simples da variável de tempo e o preditor, mas as funções mais complexas podem ser especificadas também.
- Algumas variáveis podem ter valores diferentes em períodos de tempo diferentes, mas não são sistematicamente relacionadas com o tempo. Em tais casos, é necessário definir um **preditor dependente de tempo segmentado**, que pode ser feito utilizando expressões lógicas. Expressões lógicas utilizam o valor 1 se verdadeiro e 0 se falso. Utilizando uma série de expressões lógicas, é possível criar seu preditor dependente de tempo a partir de um conjunto de medidas. Por exemplo, se você tiver a pressão sanguínea medida uma vez por semana para as quatro semanas do estudo (identificada como $BP1$ para $BP4$), será possível definir seu preditor dependente de tempo como $(T_ < 1) * BP1 + (T_ >= 1 \& T_ < 2) * BP2 + (T_ >= 2 \& T_ < 3) * BP3 + (T_ >= 3 \& T_ < 4) * BP4$. Observe que exatamente um dos termos em parênteses será igual a 1 para qualquer caso determinado e os restantes serão todos iguais a 0. Em outras palavras, esta função significa que se o tempo for menor que uma semana, use $BP1$; se for mais de uma semana, mas menos de duas semanas, use $BP2$, e assim por diante.

Nota: se o seu preditor dependente de tempo segmentado é constante dentro dos segmentos, como no exemplo da pressão sanguínea fornecido acima, pode ser mais fácil especificar as constantes em trechos, preditor dependente de tempo por assuntos de divisão entre vários casos. Consulte a discussão em Identificados de sujeito no Capítulo 12, “Regressão de Cox de amostras complexas”, na página 49 para obter mais informações.

Na caixa de diálogo Definir preditor dependente de tempo, é possível usar os controles de construção de função para construir a expressão para a covariável dependente de tempo, ou inseri-la diretamente na área de texto Expressão Numérica. Observe que as constantes de sequência devem estar entre aspas ou apóstrofos, e as constantes numéricas devem ser digitadas em formato americano, com o ponto como o delimitador decimal. A variável resultante recebe o nome que você especificar e deve ser incluída como um fator ou covariável na guia Preditores.

Subgrupos

Camada de linha de base. Um risco de base separado e a função de sobrevivência são calculados para cada valor dessa variável, enquanto um único conjunto de coeficientes do modelo é estimado entre as camadas.

Variável de subpopulação. Especifique uma variável para definir uma subpopulação. A análise é executada apenas para a categoria selecionada da variável de subpopulação.

Modelo

Especificar efeitos do modelo. Por padrão, o procedimento constrói um modelo dos principais efeitos usando os fatores e covariáveis especificados na caixa de diálogo principal. Como alternativa, é possível construir um modelo customizado que inclui efeitos de interação e termos aninhados.

Termos não aninhados

Para os fatores e covariáveis selecionados:

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

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

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

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

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

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

Termos aninhados

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

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

Limitações. Os termos aninhados possuem as restrições a seguir:

- Todos os fatores dentro de 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.

Estatísticas

Informações do plano de amostra. Exibe informações de resumo sobre a amostra, incluindo a contagem não ponderada e o tamanho da população.

Evento e sumarização de evento. Exibe informações de resumo sobre o número e a porcentagem de casos censurados.

Conjunto de risco nos tempos de evento. Exibe o número de eventos e o número de risco para cada linha de tempo de eventos em cada camada de referência.

Parâmetros. Este grupo permite controlar a exibição de estatísticas relacionadas aos parâmetros do modelo.

- **Estimativa.** Exibe estimativas dos coeficientes.

- **Estimativa exponenciada.** Exibe a base do logaritmo natural elevado à potência das estimativas dos coeficientes. Embora a estimativa tem ótimas propriedades para testes estatísticos, a estimativa exponenciada, ou $\exp(B)$, é mais fácil de interpretar.
- **Erro padrão.** Exibe o erro padrão para cada estimativa de coeficiente.
- **Intervalo de confiança.** Exibe um intervalo de confiança para cada estimativa de coeficiente. O nível de confiança para o intervalo é definido na caixa de diálogo Opções.
- **teste-t.** Exibe um teste t de cada estimativa de coeficiente. A hipótese nula para cada teste é que o valor do coeficiente é 0.
- **Covariâncias de estimativas paramétrica.** Exibe uma estimativa da matriz de covariâncias para os coeficientes do modelo.
- **Correlações de estimativas paramétrica.** Exibe uma estimativa da matriz de correlações para os coeficientes do modelo.
- **Efeito do design.** A razão da variância da estimativa para a variância obtida ao supor que a amostra é uma amostra aleatória simples. Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Isso é uma medida do efeito da especificação de um design complexo, em que os valores adicionais a partir de 1 indicam efeitos maiores.

Suposições de modelo. Este grupo permite produzir um teste da suposição de riscos proporcionais. O teste compara o modelo ajustado em um modo alternativo que inclui preditores dependentes de tempo x^*_{TF} para cada preditor x , em que $_{TF}$ é a função de tempo especificada.

- **Função de tempo.** Especifica a forma de $_{TF}$ para o modelo alternativo. Para a função **identidade**, $_{TF}=T_$. Para a função **log**, $_{TF}=\log(T_)$. Para **Kaplan-Meier**, $_{TF}=1-S_{KM}(T_)$, em que $S_{KM}(\cdot)$ é a estimativa Kaplan-Meier da função de sobrevivência. Para **ranqueamento**, $_{TF}$ é a ordem de ranqueamento de $T_$ entre os horários de encerramento observados.
- **Estimativas paramétrica para o modelo alternativo.** Exibe a estimativa, erro padrão e intervalo de confiança para cada parâmetro no modelo alternativo.
- **Matriz de covariâncias para o modelo alternativo.** Exibe a matriz de covariâncias estimadas entre parâmetros no modelo alternativo.

Sobrevivência de linha de base e funções de risco acumulativo. Exibe a função de sobrevivência da linha de base e a função de risco acumulativo, juntamente com seus erros padrão.

Nota: se preditores dependentes de tempo definidos na guia Preditores são incluídos no modelo, esta opção não estará disponível.

Gráficos

A guia Gráficos permite solicitar gráficos da função de risco, função de sobrevivência, log menos log da função de sobrevivência e um menos a função de sobrevivência. Também é possível escolher plotar os intervalos de confiança junto com as funções especificadas; o nível de confiança é definido na guia Opções.

Padrões de preditor. É possível especificar um padrão de valores do preditor a ser usado para gráficos solicitados e o arquivo de sobrevivência exportado na guia Exportar. Observe que essas opções não estarão disponíveis se preditores dependentes de tempo definidos na guia preditores forem incluídos no modelo.

- **Plotar fatores em.** Por padrão, cada fator é avaliado no seu nível mais alto. Insira ou selecione um nível diferente, se desejado. Como alternativa, é possível escolher plotar linhas separadas para cada nível de um único fator, marcando a caixa de seleção para esse fator.
- **Plotar covariáveis em.** Cada covariável é avaliada na sua média. Insira ou selecione um valor diferente se você desejar.

Testes de hipótese

Estatística do teste. Este grupo permite selecionar o tipo de estatística usada para testar hipóteses. É possível escolher entre F , F ajustado, qui-quadrado e qui-quadrado ajustado.

Amostragem de graus de liberdade. Este grupo fornece o controle sobre os graus de liberdade do plano de amostragem usados para calcular valores de p para todas as estatísticas de teste. Se baseado no plano de amostragem, o valor é a diferença entre o número de unidades de amostragem primária e o número de camadas no primeiro estágio de amostragem. Como alternativa, é possível configurar um grau de liberdade especificando um número inteiro positivo.

Ajuste para múltiplas comparações. Ao executar teste de hipóteses com diversos contrastes, o nível de significância global poderá ser ajustado a partir dos níveis de significância para os contrastes incluídos. Este grupo permite escolher o método de ajustamento.

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

Salvar o

Salvar variáveis. Este grupo permite salvar variáveis relacionadas a modelo ao conjunto para conjunto de dados ativo uso posterior em diagnósticos e relatório de resultados. Observe que nenhum destes está disponível quando preditores dependentes de tempo são incluídos no modelo.

- **Função de sobrevivência.** Salva a probabilidade de sobrevivência (o valor da função de sobrevivência) no momento observado e os valores do preditor para cada caso.
- **Limite inferior do intervalo de confiança para a função de sobrevivência.** Salva o limite inferior do intervalo de confiança para a função de sobrevivência no momento observado e os valores do preditor para cada caso.
- **Limite superior do intervalo de confiança para a função de sobrevivência.** Salva o limite superior do intervalo de confiança para a função de sobrevivência no momento observado e os valores do preditor para cada caso.
- **Função de risco acumulativo.** Salva o risco acumulativo, ou $-\ln(\text{sobrevivência})$, no momento observado e os valores do preditor para cada caso.
- **Limite inferior do intervalo de confiança para a função de risco acumulativo.** Salva o limite inferior do intervalo de confiança para a função de risco acumulativo no momento observado e os valores do preditor para cada caso.
- **Limite superior do intervalo de confiança para função de risco acumulativo.** Salva o limite superior do intervalo de confiança para a função de risco acumulativo no momento observado e os valores do preditor para cada caso.
- **Valor predito de preditor linear.** Salva a combinação linear de preditores corrigidos de valor de referência vezes coeficientes de regressão. O preditor linear é a razão da função de risco para o risco de base. No modelo de risco proporcional, esse valor é todo o tempo.
- **Resíduo de Schoenfeld.** Para cada caso sem censura e cada parâmetro não redundante no modelo, o resíduo de Schoenfeld é a diferença entre o valor observado do preditor associado com o parâmetro de modelo e o valor esperado do preditor para os casos em risco definido na hora do evento observada.

Os resíduos de Schoenfeld podem ser usados para ajudar a avaliar a suposição de riscos proporcionais; por exemplo, para um preditor x , gráficos de resíduos de Schoenfeld para o preditor dependente do tempo $x \cdot \ln(T_+)$ versus o tempo que deveria mostrar uma linha horizontal em 0 se os riscos proporcionais forem mantidos. Uma variável separada é salva para cada parâmetro não redundante no modelo. Os resíduos de Schoenfeld são calculados apenas para casos sem censura.

- **Resíduo de Martingale.** Para cada caso, o resíduo de martingale é a diferença entre a censura observada (0 se censurado, 1 se não) e a expectativa de um evento durante o tempo de observação.
- **Resíduo de deviance.** Os resíduos de deviance são resíduos de martingale "ajustados" para aparecer mais simétricos sobre 0. Os gráficos de resíduos de deviance com relação a preditores devem revelar sem padrões.
- **Resíduo de Cox-Snell.** Para cada caso, a resíduo de Cox-Snell é a expectativa de um evento durante o tempo de observação ou a censura observada menos o resíduo de martingale.
- **Resíduo de escore.** Para cada caso e cada parâmetro não redundante no modelo, o resíduo de escore é a contribuição do caso para o primeiro derivativo da pseudoverossimilhança. Uma variável separada é salva para cada parâmetro não redundante no modelo.
- **Resíduo de DFBeta.** Para cada caso e cada parâmetro não redundante no modelo, o resíduo de DFBeta aproxima a mudança no valor da estimativa paramétrica quando o caso é removido do modelo. Os casos com resíduos de DFBeta relativamente grandes podem exercer influência indevida na análise. Uma variável separada é salva para cada parâmetro não redundante no modelo.
- **Resíduos agregados.** Quando vários casos representam um único assunto, o resíduo agregado para um assunto é simplesmente a soma dos resíduos do caso correspondente sobre todos os casos que pertencem ao mesmo assunto. Para o resíduo de Schoenfeld, a versão agregada é a mesma que da versão não agregada, porque o resíduo de Schoenfeld é definido apenas para casos não censurados. Esses resíduos estão disponíveis apenas quando um assunto identificador é especificado na guia Tempo e Evento.

Nomes de variáveis salvas. A geração de nome automático assegura que você mantenha todos os seus trabalhos. Nomes customizados permitem descartar/substituir resultados de execuções anteriores sem primeiro excluir as variáveis salvas no Editor de Dados.

Exportar

Exportar modelo como dados do IBM SPSS Statistics. Grava um conjunto de dados no formato do IBM SPSS Statistics que contém a correlação paramétrica ou matriz de covariâncias com as estimativas paramétrica, erros padrão, valores de significância e graus de liberdade. A ordem das variáveis no arquivo matriz é conforme a seguir.

- **rowtype_.** Adota os valores (e rótulos de valor), COV (Covariâncias), CORR (Correlações), EST (estimativas paramétrica), SE (erros padrão), SIG (níveis de significância) e DF (graus de liberdade do design de amostragem). Há um caso separado com tipo de linha COV (ou CORR) para cada parâmetro de modelo, além de um caso separado para cada um dos outros tipos de linha.
- **varname_.** Adota os valores P1, P2,..., correspondendo a uma lista ordenada de todos os parâmetros de modelo para os tipos de linha COV ou CORR, com os rótulos de valor correspondentes às sequências paramétricas mostradas na tabela de estimativas paramétrica. As células estão em branco para outros tipos de linha.
- **P1, P2, ...** Essas variáveis correspondem a uma lista ordenada de todos os parâmetros de modelo, com os rótulos de variáveis correspondentes às sequências paramétricas mostradas na tabela de estimativas paramétrica, e adotam os valores de acordo com o tipo de linha. Para parâmetros redundantes, todas as covariâncias são definidas como zero; as correlações são configuradas como o valor omissivo do sistema; todas as estimativas paramétrica são definidas como zero; e todos os erros padrão, níveis de significância e graus de liberdade residuais são definidos como o valor omissivo do sistema.

Nota: este arquivo não é imediatamente utilizável para análises adicionais em outros procedimentos que leem um arquivo matriz, a menos que esses procedimentos aceitem todos os tipos de linhas exportados aqui.

Exportar função de sobrevivência como dados do IBM SPSS Statistics. Grava um conjunto de dados no formato do IBM SPSS Statistics, que contém a função de sobrevivência, o erro padrão da função de sobrevivência, os limites superior e inferior do intervalo de confiança da função de sobrevivência e a função de risco acumulativo para cada falha ou tempo acumulativo, avaliados na linha de base e nos padrões de preditor especificados na guia Gráfico. A ordem das variáveis no arquivo matriz é conforme a seguir.

- **Variável de camada de linha de base.** Tabelas de sobrevivência separadas são produzidas para cada valor da variável de camada.
- **Variável de tempo de sobrevivência.** O horário do evento; um caso separado é criado para cada hora do evento exclusivo.
- **Sur_0, LCL_Sur_0, UCL_Sur_0.** A função de sobrevivência da linha de base e os limites superior e inferior de seu intervalo de confiança.
- **Sur_R, LCL_Sur_R, UCL_Sur_R.** A função de sobrevivência avaliada no padrão "referência" (consulte a tabela de valores padrão na saída) e os limites superior e inferior de seu intervalo de confiança.
- **Sur_##, LCL_Sur_##, UCL_Sur_##, ...** A função de sobrevivência avaliada em cada um dos padrões de preditor especificados na guia Gráficos e os limites superior e inferior de seus intervalos de confiança. Consulte a tabela de valores padrão na saída para corresponder padrões com o número ##.
- **Haz_0, LCL_Haz_0, UCL_Haz_0.** A função de risco acumulativo de linha de base e os limites superior e inferior de seu intervalo de confiança.
- **Haz_R, LCL_Haz_R, UCL_Haz_R.** A função de risco acumulativo avaliada no padrão "referência" (consulte a tabela de valores padrão na saída) e os limites superior e inferior de seu intervalo de confiança.
- **Haz_##, LCL_Haz_##, UCL_Haz_##, ...** A função de risco acumulativo avaliada em cada um dos padrões de preditor especificados na guia Gráficos e os limites superior e inferior de seus intervalos de confiança. Consulte a tabela de valores padrão na saída para corresponder padrões com o número ##.

Exportar modelo como XML. Salva todas as informações necessárias para prever a função de sobrevivência, incluindo estimativas paramétrica e a função de sobrevivência de linha de base, 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

Estimação. Esses controles especificam critérios para estimação de coeficientes de regressão.

- **Máximo de iterações.** O número máximo de iterações que o algoritmo executará. Especifique um número inteiro não negativo.
- **Divisão máxima da etapa pela metade.** Em cada iteração, o tamanho do passo é reduzido por um fator de 0,5 até que o log da verossimilhança aumente ou a divisão máxima da etapa pela metade seja atingida. Especifique um número inteiro positivo.
- **Iterações de limite com base em mudança nas estimativas paramétrica.** Quando selecionada, o algoritmo é interrompido após uma iteração na qual uma mudança absoluta ou relativa nas estimativas paramétrica for menor que o valor especificado, que deve ser positivo.
- **Iterações de limite com base em mudança de log da verossimilhança.** Quando selecionada, o algoritmo é interrompido após uma iteração na qual uma mudança absoluta ou relativa na função de log da verossimilhança for menor que o valor especificado, que deve ser positivo.
- **Exibir histórico de iteração.** Exibe o histórico de iteração para as estimativas paramétrica e pseudo log da verossimilhança e imprime a última avaliação da mudança nas estimativas paramétrica e pseudo log da verossimilhança. A tabela de históricos de iteração imprime cada iteração n que começa com a iteração 0 (as estimativas iniciais), em que n é o valor do incremento. Se o histórico de iteração for solicitado, então a última iteração será sempre exibida, independentemente de n .

- **Método de quebra de empate para estimação paramétrica.** Quando houver tempos de falha empatados observados, um desses métodos será usado para quebrar os empates. O método Efron é mais caro em termos computacionais.

Funções de sobrevivência. Esses controle especificam critérios para cálculos que envolvem a função de sobrevivência.

- **Método para estimar funções de sobrevivência de linha de base.** O método **Breslow** (Nelson-Aalan ou empírico) estima o risco acumulativo de linha de base por uma função de etapa não decrescente com etapas nos tempos de falha observados, em seguida, calcula a sobrevivência de linha de base pela relação $\text{survival}=\exp(\text{risco acumulativo})$. O método de **Efron** é mais computacionalmente caro e é reduzido para o método de Breslow quando não há empates. O método de **limite de produto** estima a sobrevivência de linha de base por uma função contínua da direita não crescente; quando não há preditores no modelo, esse método é reduzido para a estimação de Kaplan-Meier.
- **Intervalos de confiança de funções de sobrevivência.** O intervalo de confiança pode ser calculado de três maneiras: em unidades originais, por meio de uma transformação de log ou em uma transformação de log menos log. Apenas a transformação de log menos log garante que os limites do intervalo de confiança serão mantidos entre 0 e 1, mas a transformação logarítmica geralmente parece executar "melhor".

Valores omissos do usuário. Todas as variáveis devem ter valores válidos para um caso a serem incluídos na análise. Estes controles permitem decidir se valores omissos de usuário são tratados como válidos entre modelos categórico (incluindo fatores, evento, camadas e variáveis de subpopulação) e variáveis de plano de amostragem.

Intervalo de confiança(%). Este é o nível de intervalo de confiança usado para estimativas de coeficiente, estimativas de coeficientes exponenciados, estimativas de função de sobrevivência e estimativas de função de risco acumulativo. Especifique um valor maior ou igual a 0 e menor que 100.

Recursos adicionais de comando CSCOXREG

O idioma do comando também permite:

- Executar teste de hipótese ccustomizadas (usando o subcomando CUSTOM e /PRINT LMATRIX).
- Especificação de tolerância (usando /CRITERIA SINGULAR).
- Tabela de funções estimáveis gerais (usando o /PRINT GEF).
- Padrões preditores múltiplos (usando submocandos PATTERN múltiplos).
- Número máximo de variáveis salvas quando um nome raiz é especificado (usando o subcomando SAVE). O diálogo respeita o padrão CSCOXREG de 25 variáveis.

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

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

amostra
 design complexo 5
amostragem aleatória simples
 no Assistente de amostragem 6
amostragem complexa
 plano de amostra 5
 plano de análise 13
amostragem PPS
 no Assistente de amostragem 6
amostragem sequencial
 no Assistente de amostragem 6
amostragem sistemática
 no Assistente de amostragem 6
Amostras complexas
 opções 20, 24, 27, 30
 testes de hipótese 33, 39, 45
 valores omissos 20, 26
arquivo de plano 2

B

Bonferroni
 em Amostras Complexas 33, 39, 45
 na Regressão de Cox de amostras complexas 54

C

camada de linha de base
 na Regressão de Cox de amostras complexas 51
categoria de referência
 na Regressão logística de amostras complexas 37
 no Modelo linear geral de amostras complexas 33
categorias preditas
 na Regressão logística de amostras complexas 40
 na Regressão ordinal de amostras complexas 46
clusters
 no Assistente de amostragem 5
 no Assistente de preparação de análise 13
coeficiente de variação (COV)
 em Amostras complexas descritivas 23
 em Frequências de amostras complexas 19
 em Razões de amostras complexas 29
 nas Tabulações cruzadas de amostras complexas 25
contagem não ponderada
 em Amostras complexas descritivas 23
 em Frequências de amostras complexas 19

contagem não ponderada (*continuação*)
 em Razões de amostras complexas 29
 nas Tabulações cruzadas de amostras complexas 25
contrastes
 no Modelo linear geral de amostras complexas 33
contrastes de desvio
 no Modelo linear geral de amostras complexas 33
contrastes de diferença
 no Modelo linear geral de amostras complexas 33
contrastes de Helmert
 no Modelo linear geral de amostras complexas 33
contrastes polinomiais
 no Modelo linear geral de amostras complexas 33
contrastes repetidos
 no Modelo linear geral de amostras complexas 33
contrastes simples
 no Modelo linear geral de amostras complexas 33
convergência de probabilidade
 na Regressão logística de amostras complexas 40
 na Regressão ordinal de amostras complexas 47
convergência paramétrica
 na Regressão logística de amostras complexas 40
 na Regressão ordinal de amostras complexas 47
correção sequencial de Bonferroni
 em Amostras Complexas 33, 39, 45
 na Regressão de Cox de amostras complexas 54
correção Sidak
 em Amostras Complexas 33, 39, 45
 na Regressão de Cox de amostras complexas 54
correção Sidak sequencial
 em Amostras Complexas 33, 39, 45
 na Regressão de Cox de amostras complexas 54
correlações de estimativas paramétrica
 na Regressão logística de amostras complexas 38
 na Regressão ordinal de amostras complexas 44
 no Modelo linear geral de amostras complexas 32
covariâncias de estimativas paramétrica
 na Regressão logística de amostras complexas 38
 na Regressão ordinal de amostras complexas 44

covariâncias de estimativas paramétrica (*continuação*)
 no Modelo linear geral de amostras complexas 32

D

Descrições de amostras complexas 23
 estatísticas 23
 valores omissos 24
diferença de risco
 nas Tabulações cruzadas de amostras complexas 25
diferença menos significativa
 em Amostras Complexas 33, 39, 45
 na Regressão de Cox de amostras complexas 54

E

efeito do design
 em Amostras complexas descritivas 23
 em Frequências de amostras complexas 19
 em Razões de amostras complexas 29
 na Regressão de Cox de amostras complexas 52
 na Regressão logística de amostras complexas 38
 na Regressão ordinal de amostras complexas 44
 nas Tabulações cruzadas de amostras complexas 25
 no Modelo linear geral de amostras complexas 32
efeito do design da raiz quadrada
 em Frequências de amostras complexas 19
 na Regressão de Cox de amostras complexas 52
erro padrão
 em Amostras complexas descritivas 23
 em Frequências de amostras complexas 19
 em Razões de amostras complexas 29
 na Regressão logística de amostras complexas 38
 na Regressão ordinal de amostras complexas 44
 nas Tabulações cruzadas de amostras complexas 25
 no Modelo linear geral de amostras complexas 32
estatística de F
 em Amostras Complexas 33, 39, 45

- estatística de F (*continuação*)
 - na Regressão de Cox de amostras complexas 54
- estatística F ajustada
 - em Amostras Complexas 33, 39, 45
 - na Regressão de Cox de amostras complexas 54
- estatística R2
 - no Modelo linear geral de amostras complexas 32
- estimação de amostragem
 - no Assistente de preparação de análise 14
- estimativas paramétrica
 - na Regressão de Cox de amostras complexas 52
 - na Regressão logística de amostras complexas 38
 - na Regressão ordinal de amostras complexas 44
 - no Modelo linear geral de amostras complexas 32
- estratificação
 - no Assistente de amostragem 5
 - no Assistente de preparação de análise 13
- etapa pela metade
 - na Regressão logística de amostras complexas 40
 - na Regressão ordinal de amostras complexas 47

F

- Frequências de amostras complexas 19
- estatísticas 19

G

- graus de liberdade
 - em Amostras Complexas 33, 39, 45
 - na Regressão de Cox de amostras complexas 54

H

- histórico de iteração
 - na Regressão logística de amostras complexas 40
 - na Regressão ordinal de amostras complexas 47

I

- informações de plano de amostra
 - na Regressão de Cox de amostras complexas 52
- intervalos de confiança
 - em Amostras complexas descritivas 23
 - em Frequências de amostras complexas 19
 - em Razões de amostras complexas 29

- intervalos de confiança (*continuação*)
 - na Regressão logística de amostras complexas 38
 - na Regressão ordinal de amostras complexas 44
 - nas Tabulações cruzadas de amostras complexas 25
 - no Modelo linear geral de amostras complexas 32, 34
- iterações
 - na Regressão logística de amostras complexas 40
 - na Regressão ordinal de amostras complexas 47

M

- média
 - em Amostras complexas descritivas 23
- médias marginais estimadas
 - no Modelo linear geral de amostras complexas 33
- medida de tamanho
 - no Assistente de amostragem 6
- método de amostragem
 - no Assistente de amostragem 6
- método de amostragem de Brewer
 - no Assistente de amostragem 6
- método de amostragem de Murthy
 - no Assistente de amostragem 6
- método de amostragem de Sampford
 - no Assistente de amostragem 6
- Método de estimação de Breslow
 - na Regressão de Cox de amostras complexas 56
- Método de estimação de Efron
 - na Regressão de Cox de amostras complexas 56
- método Newton-Raphson
 - na Regressão ordinal de amostras complexas 47
- Modelo linear geral de amostras complexas 31
 - estatísticas 32
 - médias estimadas 33
 - model 31
 - opções 34
 - recursos adicionais do comando 35
 - salvar variáveis 34

N

- nível de confiança
 - na Regressão logística de amostras complexas 40
 - na Regressão ordinal de amostras complexas 47

P

- plano de amostra 5
- plano de análise 13
- ponderações da amostra
 - no Assistente de amostragem 8

- ponderações de amostra de entrada
 - no Assistente de amostragem 5
- Pontuação de Fisher
 - na Regressão ordinal de amostras complexas 47
- porcentagens da tabela
 - em Frequências de amostras complexas 19
 - nas Tabulações cruzadas de amostras complexas 25
- porcentagens de coluna
 - nas Tabulações cruzadas de amostras complexas 25
- porcentagens de linha
 - nas Tabulações cruzadas de amostras complexas 25
- preditor dependente de tempo
 - na Regressão de Cox de amostras complexas 51
- probabilidade predita
 - na Regressão logística de amostras complexas 40
 - na Regressão ordinal de amostras complexas 46
- probabilidades acumulativas
 - na Regressão ordinal de amostras complexas 46
- probabilidades de inclusão
 - no Assistente de amostragem 8
- probabilidades de resposta
 - na Regressão ordinal de amostras complexas 44
- proporção de amostra
 - no Assistente de amostragem 8
- pseudoestatísticas R2
 - na Regressão logística de amostras complexas 38
 - na Regressão ordinal de amostras complexas 44

Q

- qui-quadrado
 - em Amostras Complexas 33, 39, 45
 - na Regressão de Cox de amostras complexas 54
- qui-quadrado ajustado
 - em Amostras Complexas 33, 39, 45
 - na Regressão de Cox de amostras complexas 54

R

- raiz quadrada de efeito do design
 - em Amostras complexas descritivas 23
- em Razões de amostras complexas 29
- na Regressão logística de amostras complexas 38
- na Regressão ordinal de amostras complexas 44
- nas Tabulações cruzadas de amostras complexas 25
- no Modelo linear geral de amostras complexas 32

- Razões de amostras complexas 29
 - estatísticas 29
 - valores omissos 30
 - razões de chances
 - na Regressão logística de amostras complexas 40
 - na Regressão ordinal de amostras complexas 46
 - nas Tabulações cruzadas de amostras complexas 25
 - Regressão de Cox de amostras complexas
 - análise de Kaplan-Meier 49
 - definir evento 50
 - estatísticas 52
 - exportar modelo 55
 - gráficos 53
 - model 52
 - opções 56
 - preditor dependente de tempo 51
 - preditores 51
 - salvar variáveis 54
 - subgrupos 51
 - testes de hipótese 54
 - variáveis de data e hora 49
 - Regressão logística de amostras complexas 37
 - categoria de referência 37
 - estatísticas 38
 - model 38
 - opções 40
 - razões de chances 40
 - recursos adicionais do comando 41
 - salvar variáveis 40
 - Regressão ordinal de amostras complexas 43
 - estatísticas 44
 - model 44
 - opções 47
 - probabilidades de resposta 44
 - razões de chances 46
 - salvar variáveis 46
 - resíduos
 - nas Tabulações cruzadas de amostras complexas 25
 - no Modelo linear geral de amostras complexas 34
 - resíduos agregados
 - na Regressão de Cox de amostras complexas 54
 - resíduos ajustados
 - nas Tabulações cruzadas de amostras complexas 25
 - resíduos de Cox-Snell
 - na Regressão de Cox de amostras complexas 54
 - resíduos de deviance
 - na Regressão de Cox de amostras complexas 54
 - resíduos de escore
 - na Regressão de Cox de amostras complexas 54
 - resíduos de martingale
 - na Regressão de Cox de amostras complexas 54
 - Resíduos parciais de Schoenfeld
 - na Regressão de Cox de amostras complexas 54
 - risco relativo
 - nas Tabulações cruzadas de amostras complexas 25
- S**
- SampleWeights
 - no Assistente de preparação de análise 13
 - separação
 - na Regressão logística de amostras complexas 40
 - na Regressão ordinal de amostras complexas 47
 - soma
 - em Amostras complexas descritivas 23
 - subpopulação
 - na Regressão de Cox de amostras complexas 51
- T**
- tabelas de classificação
 - na Regressão logística de amostras complexas 38
 - na Regressão ordinal de amostras complexas 44
 - Tabulações cruzadas de amostras complexas 25
 - estatísticas 25
 - tamanho da amostra
 - no Assistente de amostragem 7, 8
 - tamanho da população
 - em Amostras complexas descritivas 23
 - em Frequências de amostras complexas 19
 - em Razões de amostras complexas 29
 - nas Tabulações cruzadas de amostras complexas 25
 - no Assistente de amostragem 8
 - teste de linhas paralelas
 - na Regressão ordinal de amostras complexas 44
 - teste de riscos proporcionais
 - na Regressão de Cox de amostras complexas 52
 - teste t
 - na Regressão logística de amostras complexas 38
 - na Regressão ordinal de amostras complexas 44
 - no Modelo linear geral de amostras complexas 32
- V**
- valores acumulativos
 - em Frequências de amostras complexas 19
 - valores esperados
 - nas Tabulações cruzadas de amostras complexas 25
 - valores omissos
 - em Amostras Complexas 20, 26
 - em Amostras complexas descritivas 24
 - em Razões de amostras complexas 30
 - na Regressão logística de amostras complexas 40
 - na Regressão ordinal de amostras complexas 47
 - no Modelo linear geral de amostras complexas 34
 - valores preditos
 - no Modelo linear geral de amostras complexas 34



Impresso no Brasil