

IBM SPSS Complex Samples 29



Nota

Antes de usar estas informações e o produto suportado por elas, leia as informações nos [“Avisos”](#) na [página 41](#).

Informações do produto

Esta edição aplica-se à versão 29, 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.

© **Copyright International Business Machines Corporation .**

Capítulo 1. Amostras complexas.....	1
Introdução a procedimentos de amostras complexas.....	1
Propriedades das amostras complexas.....	1
Uso de procedimentos de amostras complexas.....	2
Leituras adicionais.....	2
Amostragem de um design complexo.....	3
Criando um novo plano de amostra.....	3
Assistente de amostragem: variáveis de design.....	3
Assistente de amostragem: Método de amostragem.....	4
Assistente de amostragem: tamanho da amostra.....	5
Assistente de amostragem: variáveis de saída.....	6
Assistente de amostragem: Sumarização do plano.....	6
Assistente de amostragem: desenhar opções de seleção de amostras.....	6
Assistente de amostragem: desenhar arquivos de saída de amostra.....	6
Assistente de amostragem: Concluir.....	7
Modificando um plano de amostra existente.....	7
Assistente de amostragem: Sumarização do plano.....	7
Executando um plano de amostra existente.....	7
Recursos adicionais de comandos CSPLAN e CSSELECT.....	8
Preparando uma amostra complexa para análise.....	8
Criando um novo plano de análise.....	8
Assistente de preparação de análise: variáveis de design.....	9
Assistente de preparação de análise: método de estimação.....	9
Assistente de preparação de análise: tamanho.....	9
Assistente de preparação de análise: resumo do plano.....	10
Assistente de preparação de análise: conclusão.....	11
Modificando um plano de análise existente.....	11
Assistente de preparação de análise: resumo do plano.....	11
Plano de amostras complexas.....	11
Frequências de amostras complexas.....	11
Estatísticas de frequências de amostras complexas.....	12
Valores omissos de amostras complexas.....	13
Opções de amostras complexas.....	13
Descrições de amostras complexas.....	13
Estatísticas descritivas de amostras complexas.....	14
Valores omissos de amostras complexas descritivas.....	14
Opções de amostras complexas.....	15
Tabulações cruzadas de amostras complexas.....	15
Estatísticas de Tabulações cruzadas de amostras complexas.....	15
Valores omissos de amostras complexas.....	16
Opções de amostras complexas.....	17
Razões de amostras complexas.....	17
Estatísticas de razão de amostras complexas.....	18
Valores omissos de razões de amostras complexas.....	18
Opções de amostras complexas.....	18
Modelo linear geral de amostras complexas.....	18
Modelo linear geral de amostras complexas.....	19
Estatísticas de modelo linear geral de amostras complexas.....	20
Testes de hipótese de amostras complexas.....	20
Médias estimadas do modelo linear geral de amostras complexas.....	21
Salvar modelo linear geral de amostras complexas.....	21

Opções de modelo linear geral de amostras complexas.....	22
Recursos adicionais de comando CSGLM.....	22
Regressão logística de amostras complexas.....	22
Categoria de referência de regressão logística de amostras complexas.....	23
Modelo de regressão logística de Amostras Complexas.....	23
Estatísticas de Regressão logística de amostras complexas.....	24
Testes de hipótese de amostras complexas.....	24
Razões de chances de regressão logística de amostras complexas.....	25
Regressão logística de amostras complexas - Salvar.....	25
Opções de Regressão logística de amostras complexas.....	26
Recursos adicionais do comando CSLOGISTIC.....	26
Regressão ordinal de amostras complexas.....	27
Probabilidades de resposta de regressão ordinal de amostras complexas.....	27
Modelo de regressão ordinal de amostras complexas.....	28
Estatísticas de regressão ordinal de amostras complexas.....	28
Testes de hipótese de amostras complexas.....	29
Razões de chances de regressão ordinal de amostras complexas.....	30
Salvar Regressão Ordinal de Amostras Complexas.....	30
Opções de regressão ordinal de amostras complexas.....	30
Recursos adicionais de comando CSORDINAL.....	31
Regressão de Cox de amostras complexas.....	31
Definir evento	33
Preditores	33
Subgrupos	34
Modelo	34
Estatísticas	35
Gráficos	36
Testes de hipótese	36
Salvar	37
Exportar	38
Opções	39
Recursos adicionais de comando CSCOXREG.....	39
Avisos.....	41
Marcas comerciais.....	42
Índice remissivo.....	45

Capítulo 1. Amostras complexas

Os recursos de amostras complexas a seguir estão incluídos em SPSS Statistics Premium Edition ou a opção Amostras complexas.

Introdução a procedimentos de amostras complexas

Uma suposição inerente aos 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 consideram econômico e conveniente obter amostras de maneira mais estruturada.

A opção Amostras complexas permite selecionar uma amostra de acordo com um projeto complexo e incorporar as especificações do projeto na análise de dados, garantindo assim 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 amostrais individuais são selecionadas aleatoriamente com igual probabilidade e sem reposição (WOR) diretamente de toda a população. Por outro lado, uma determinada amostra complexa pode ter alguns ou todos os recursos a seguir:

Estratificação. A amostragem estratificada envolve a seleção de amostras independentemente dentro de subgrupos não sobrepostos da população, ou camadas. Por exemplo, os estratos podem ser grupos socioeconômicos, categorias de trabalho, faixas etárias ou grupos étnicos. Com a estratificação, é possível assegurar tamanhos de amostra adequados para subgrupos de interesse, melhorar a precisão das estimativas gerais e usar diferentes métodos de amostragem de camada a camada.

Armazenamento em cluster. A amostragem por conglomerados envolve a seleção de grupos de unidades de amostragem ou clusters. Por exemplo, os conglomerados 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 vários estágios, você seleciona uma amostra de primeiro estágio com base em clusters. Em seguida, você cria uma amostra de segundo estágio desenhando subamostras dos clusters selecionados. Se a amostra do segundo estágio for baseada em subclusters, será possível incluir um terceiro estágio na amostra. Por exemplo, na primeira etapa de uma pesquisa de opinião, uma amostra de cidades pode ser desenhada. Em seguida, a partir das cidades selecionadas, as famílias podem ser amostradas. Finalmente, dos domicílios selecionados, os indivíduos poderiam ser pesquisados. Os assistentes de Preparação de Amostragem e Análise permitem especificar três estágios em um design.

Amostragem não aleatória. Quando a seleção aleatória é difícil de obter, as unidades podem ser amostradas sistematicamente (em um intervalo fixo) ou sequencialmente.

Probabilidades de seleção diferentes. Ao amostrar clusters que contêm números de unidades desiguais, é 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. A amostragem de PPS também pode 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. Os pesos amostrais são calculados automaticamente ao desenhar uma amostra complexa e correspondem idealmente à "frequência" que cada unidade amostral 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 pesos de amostragem para analisar adequadamente uma amostra complexa. Observe que esses pesos devem ser usados

inteiramente na opção Amostras Complexas e não devem ser usados com outros procedimentos analíticos por meio do procedimento Peso de Casos, que trata os pesos como replicações de casos.

Uso de procedimentos de amostras complexas

Seu uso de procedimentos de amostras complexas depende de suas necessidades específicas. Os tipos de usuários primários são aqueles que:

- Planejam e realizam pesquisas de acordo com desenhos complexos, possivelmente analisando a amostra posteriormente. A principal ferramenta para os pesquisadores é o [Assistente de amostragem](#).
- Analisam arquivos de dados de amostra previamente obtidos de acordo com projetos complexos. Antes de usar os procedimentos de análise de Amostras Complexas, pode ser necessário usar o [Assistente de preparação de análise](#).

Independentemente do tipo de usuário que você é, é necessário fornecer informações de projeto 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. Existem dois tipos de arquivos de plano:

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

Plano de análise. Este arquivo de plano contém informações necessárias para os procedimentos de análise de amostras complexas para calcular corretamente as estimativas de variação para uma amostra complexa. O plano inclui a estrutura da amostra, métodos de estimativa para cada estágio e referências às variáveis necessárias, como pesos da amostra. O Assistente de Preparação de Análise permite criar e editar planos de análise.

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

- Um agrimensor pode especificar o primeiro estágio de um plano de amostragem de vários estágios e desenhar as unidades do primeiro estágio agora, coletar informações sobre as unidades de amostragem para o segundo estágio e, em seguida, modificar o plano de amostragem para incluir o segundo estágio.
- Um analista que não tem acesso ao arquivo do plano de amostragem pode especificar um plano de análise e fazer referência a esse plano a partir de cada procedimento de análise de Amostras Complexas.
- Um designer de amostras de uso público em larga escala pode publicar o arquivo de plano de amostragem, o 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 textos a seguir:

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.

Amostragem de um design complexo

O Assistente de amostragem orienta você pelas etapas para criar, modificar ou executar um arquivo de plano de amostragem. Antes de usar o assistente, deve-se 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. A partir dos 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 pelo assistente.

4. Opcionalmente, na etapa Variáveis de design, é possível definir estratos, clusters e pesos de amostra de entrada. Após defini-los, clique em **Avançar**.

5. Opcionalmente, na etapa 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. Na etapa Tamanho da amostra, especifique o número ou a proporção de unidades a serem amostradas.

7. Agora é possível clicar em **Concluir** para exibir a amostra.

Opcionalmente, em outras etapas, é possível:

- Escolha as variáveis de saída para salvar.
- Inclua um segundo ou terceiro estágio no design.
- Configurar várias opções de seleção, incluindo de quais estágios extrair amostras, a semente do número aleatório e se os valores omissos do usuário devem ser tratados como valores válidos das variáveis de design.
- Escolha onde salvar os dados de saída.
- Cole suas seleções como sintaxe de comando.

Assistente de amostragem: variáveis de design

Esta etapa permite selecionar variáveis de estratificação e armazenamento em cluster e definir os pesos da amostra de entrada. Também é possível especificar um rótulo para o estágio.

Estratificar por. A classificação cruzada de variáveis de estratificação define subpopulações distintas, ou estratos. Amostras separadas são obtidas para cada estrato. Para melhorar a precisão de suas estimativas, as unidades dentro dos estratos devem ser o mais homogêneas possível para as características de interesse.

Clusters. As variáveis de cluster definem grupos de unidades observacionais, ou clusters. Os agrupamentos são úteis quando a amostragem direta de unidades observacionais da população é cara ou impossível; em vez disso, você pode amostrar clusters da população e, em seguida, amostrar unidades observacionais dos clusters selecionados. No entanto, o uso de conglomerados pode introduzir correlações entre as unidades amostrais, resultando em perda de precisão. Para minimizar esse efeito, as unidades dentro dos 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” na página 4 para obter mais informações

Ponderação de amostra de entrada. Se o desenho de amostra atual for parte de um desenho de amostra maior, você pode ter pesos de amostra de um estágio anterior do maior design. É possível especificar uma

variável numérica contendo esses pesos no primeiro estágio do design atual. As ponderações de amostra são calculadas automaticamente para os 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 estágio.

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 uma etapa específica são removidas da lista em todas as etapas. As variáveis retornadas à lista de origem aparecem na lista em todas as etapas.

Controles de árvore para navegação no Assistente de amostragem

No lado esquerdo de cada etapa do assistente de amostragem há um resumo de todas as etapas. É possível navegar no assistente clicando no nome de uma etapa ativada na estrutura de tópicos. As etapas são ativadas desde que todas as etapas anteriores sejam válidas, ou seja, se cada etapa anterior tiver recebido as especificações mínimas exigidas para essa etapa. Consulte a Ajuda para etapas individuais para obter mais informações sobre por que uma determinada etapa pode ser inválida.

Assistente de amostragem: Método de amostragem

Esta etapa 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 que você escolha se deseja amostrar com reposição (WR) ou sem reposição (WOR). Consulte as descrições dos tipos para obter mais informações. Observe que alguns tipos de probabilidade proporcional ao tamanho (PPS) estão disponíveis apenas quando os clusters foram definidos e que todos os tipos de PPS estão disponíveis apenas no primeiro estágio de um projeto. Além disso, os métodos WR estão disponíveis apenas no último estágio de um design.

- **Amostragem aleatória simples.** As unidades são selecionadas com probabilidade igual. Eles podem ser selecionados com ou sem substituição.
- **Sistemático simples.** As unidades são selecionadas em um intervalo fixo ao longo do quadro de amostragem (ou camadas, se tiverem sido especificados) e extraídas sem substituição. Uma unidade selecionada aleatoriamente dentro do primeiro intervalo é escolhida como ponto de partida.
- **Sequencial simples.** As unidades são selecionadas sequencialmente com igual probabilidade e sem substituição.
- **PPS.** Este é um método de primeiro estágio que seleciona unidades aleatoriamente com probabilidade proporcional ao tamanho. Quaisquer unidades podem ser selecionadas com substituição; apenas clusters podem ser amostrados sem substituição.
- **PPS sistemático.** Este é um método de primeiro estágio que seleciona sistematicamente unidades com probabilidade proporcional ao tamanho. Eles são selecionados sem substituição.
- **PPS sequencial.** Este é um método de primeiro estágio que seleciona unidades sequencialmente com probabilidade proporcional ao tamanho do cluster e sem substituição.
- **PPS Brewer.** Este é um método de primeiro estágio que seleciona dois clusters de cada estrato com probabilidade proporcional ao tamanho do cluster e sem substituição. Uma variável de cluster deve ser especificada para usar esse método.
- **PPS Murthy.** Este é um método de primeiro estágio que seleciona dois clusters de cada estrato com probabilidade proporcional ao tamanho do cluster e sem substituição. Uma variável de cluster deve ser especificada para usar esse método.
- **PPS Sampford.** Este é um método de primeiro estágio que seleciona mais de dois clusters de cada estrato 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 esse método.
- **Usar estimativa WR para análise.** Por padrão, um método de estimativa é especificado no arquivo de plano que é consistente com o método de amostragem selecionado. Isso permite que você use a estimativa com reposição mesmo que o método de amostragem implique a estimativa 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 inferiores e superiores no MOS, substituindo quaisquer 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

Esta etapa permite especificar o número ou a proporção de unidades a serem amostradas no estágio atual. O tamanho da amostra pode ser fixo ou pode variar entre os estratos. Com o objetivo de especificar o tamanho da amostra, os conglomerados escolhidos nas etapas anteriores podem ser usados para definir os estratos.

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

- **Valor.** Um único valor é aplicado a todos os estratos. 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. Salvo amostragem com reposição, os valores de proporção também não devem ser maiores que 1.
- **Valores desiguais para camadas.** Permite inserir valores de tamanho 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 estratos.

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

Definir tamanhos desiguais

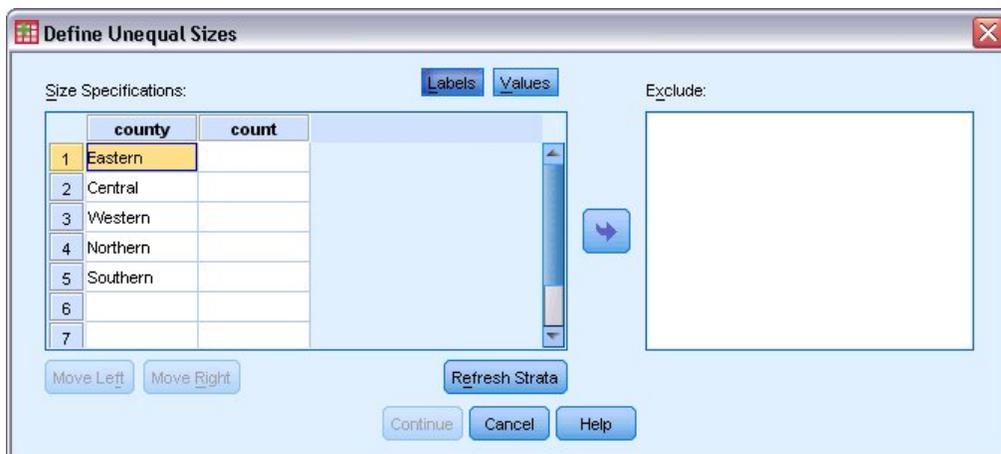


Figura 1. Caixa de diálogo Definir tamanhos desiguais

A caixa de diálogo Definir tamanhos desiguais permite inserir tamanhos por estrato.

Grade de especificações de tamanho. A grade exibe as classificações cruzadas de até cinco variáveis de estratos ou clusters — uma combinação de estratos/cluster por linha. As variáveis de grade elegíveis incluem todas as variáveis de estratificação dos estágios atuais e anteriores e todas as variáveis de cluster dos estágios anteriores. As variáveis podem ser reordenadas dentro da grade ou movidas para a lista de 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.

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

Assistente de amostragem: variáveis de saída

Esta etapa permite que você escolha variáveis para salvar quando a amostra for retirada.

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. Elas incluem:

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

Ponderação acumulativa. O peso cumulativo da amostra sobre os estágios anteriores 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

Esta é a última etapa de cada estágio, fornecendo um resumo das especificações do plano de amostra até o estágio atual. A partir daqui, é possível prosseguir para o próximo estágio (criando-o, se necessário) ou configurar opções para desenhar a amostra.

Assistente de amostragem: desenhar opções de seleção de amostras

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

Desenhar amostra. Além de escolher se deseja extrair uma amostra, também é possível optar por executar parte do desenho de amostragem. Os estágios devem ser sorteados em ordem, ou seja, o estágio 2 não pode ser sorteado a menos que o estágio 1 também seja sorteado. Ao editar ou executar um plano, não é possível testar novamente estágios bloqueados.

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

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

Dados já ordenados. Se o seu quadro de amostra for pré-classificado pelos valores das variáveis de estratificação, esta opção permite acelerar o processo de seleção.

Assistente de amostragem: desenhar arquivos de saída de amostra

Esta etapa permite que você escolha para onde direcionar casos amostrados, variáveis de peso, probabilidades conjuntas e regras de seleção de casos.

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 estão disponíveis nas sessões subsequentes, a menos que você os salve explicitamente como arquivos de dados. Os nomes dos conjuntos de dados devem seguir as regras de nomenclatura de variáveis. Se um arquivo externo ou um novo conjunto de dados for especificado, as variáveis de saída de amostragem e as variáveis no conjunto de dados ativo para os casos selecionados serão gravadas.

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 são produzidas se o método PPS WOR, PPS Brewer, PPS Sampford ou PPS Murthy for selecionado e a estimativa WR não for especificada.

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

Assistente de amostragem: Concluir

Esta é a etapa final. É possível salvar o arquivo de plano e desenhar 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 substituir o arquivo existente. Ao incluir estágios sem fazer mudanças nos estágios existentes, o Assistente substituiu 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. A partir dos 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 pelo assistente.

4. Revise o plano de amostragem no passo Sumarização do plano e, em seguida, clique em **Avançar**.

As etapas subsequentes são basicamente as mesmas de um novo design. Consulte a Ajuda para passos individuais para obter informações adicionais.

5. Navegue até a etapa Concluir e especifique um novo nome para o arquivo de plano editado ou opte por substituir o arquivo de plano existente.

Como alternativa, é possível:

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

Assistente de amostragem: Sumarização do plano

Esta etapa permite revisar o plano de amostragem e indicar os estágios que já foram amostrados. Se estiver editando um plano, você também poderá remover estágios do plano.

Estágios amostrados anteriormente. Se um quadro de amostragem estendido não estiver disponível, você terá que executar um design de amostragem multiestágio um estágio de cada vez. Selecione quais estágios já foram amostrados na lista suspensa. Quaisquer estágios que foram executados são bloqueados; eles não estão disponíveis na etapa Desenhar Opções de Seleção de Amostra e não podem ser alterados ao editar um plano.

Remover estágios. Você pode remover os estágios 2 e 3 de um design de vários estágios.

Executando um plano de amostra existente

1. A partir dos 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 pelo assistente.
 4. Revise o plano de amostragem no passo Sumarização do plano e, em seguida, clique em **Avançar**.
 5. As etapas individuais que contêm informações de estágio são ignoradas ao executar um plano de amostra. Agora é possível ir para a etapa 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 o resumo por etapas do plano que é exibido se uma amostra for projetada ou modificada, suprimir o resumo da distribuição de casos amostrados por estratos que é mostrado se o design da amostra for executado e solicitar um resumo 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.

Preparando uma amostra complexa para análise

O Assistente de Preparação de Análise orienta você pelas 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, você deve ter uma amostra desenhada de acordo com um design complexo.

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

Criando um novo plano de análise

1. A partir dos 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 os pesos de amostra na etapa Variáveis de design, opcionalmente definindo estratos e clusters.
5. É possível agora clicar em **Concluir** para salvar o plano.

Opcionalmente, em outras etapas, é possível:

- Selecione o método para estimar erros padrão na etapa Método de estimativa.
- Especifique o número de unidades amostradas ou a probabilidade de inclusão por unidade na etapa Tamanho.
- Inclua um segundo ou terceiro estágio no design.
- Cole suas seleções como sintaxe de comando.

Assistente de preparação de análise: variáveis de design

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

camada. A classificação cruzada de variáveis de estratificação define subpopulações distintas, ou estratos. Sua amostra total representa a combinação de amostras independentes de cada camada.

Clusters. As variáveis de cluster definem grupos de unidades observacionais, ou clusters. Amostras extraídas em vários estágios selecionam clusters nos estágios anteriores e, em seguida, subamostra unidades dos clusters selecionados. Ao analisar um arquivo de dados obtido por amostragem de clusters com reposição, deve-se incluir o índice de duplicação como variável de cluster.

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

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 estágio.

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 à lista de origem aparecem em todas as etapas.

Controles de árvore para navegação no Assistente de análise

No lado esquerdo de cada etapa do Assistente de análise há um resumo de todas as etapas. É possível navegar no assistente clicando no nome de uma etapa ativada na estrutura de tópicos. As etapas são ativadas desde que todas as etapas anteriores sejam válidas, ou seja, desde que cada etapa anterior tenha recebido as especificações mínimas exigidas para essa etapa. Para obter mais informações sobre por que uma determinada 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 estimativa para o estágio.

WR (amostragem com substituição). A estimativa WR não inclui uma correção para amostragem de um preenchimento finito (FPC) ao estimar a variância sob o projeto de amostragem complexo. É possível optar por incluir ou excluir o FPC ao estimar a variância sob amostragem aleatória simples (SRS).

A escolha de não incluir o FPC para estimativa de variância SRS é recomendada quando os pesos da análise foram dimensionados para que não correspondam ao tamanho da população. A estimativa de variância SRS é usada em estatísticas de computação como o efeito de design. A estimativa de WR pode ser especificada apenas no estágio final de um projeto; o assistente não permitirá que você inclua outro estágio se você selecionar a estimativa de WR.

WOR igual (amostragem de probabilidade igual sem substituição). A estimativa de Equal WOR inclui a correção de 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 de população finita, Unequal WOR contabiliza unidades de amostragem (geralmente conglomerados) selecionadas com probabilidade desigual. Este método de estimação está disponível apenas no primeiro estágio.

Assistente de preparação de análise: tamanho

Esta etapa é usada para especificar probabilidades de inclusão ou tamanhos de população para o estágio atual. Os tamanhos podem ser fixos ou podem variar entre os estratos. Com a finalidade de especificar tamanhos, os clusters especificados nas etapas anteriores podem ser usados para definir estratos. Observe que esta etapa é necessária apenas quando Equal WOR é escolhido como o Método de Estimativa.

Unidades. É possível especificar os tamanhos exatos da população ou as probabilidades com as quais as unidades foram amostradas.

- **Valor.** Um único valor é aplicado a todos os estratos. 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 desiguais para camadas.** Permite inserir valores de tamanho 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 estratos.

Definir tamanhos desiguais

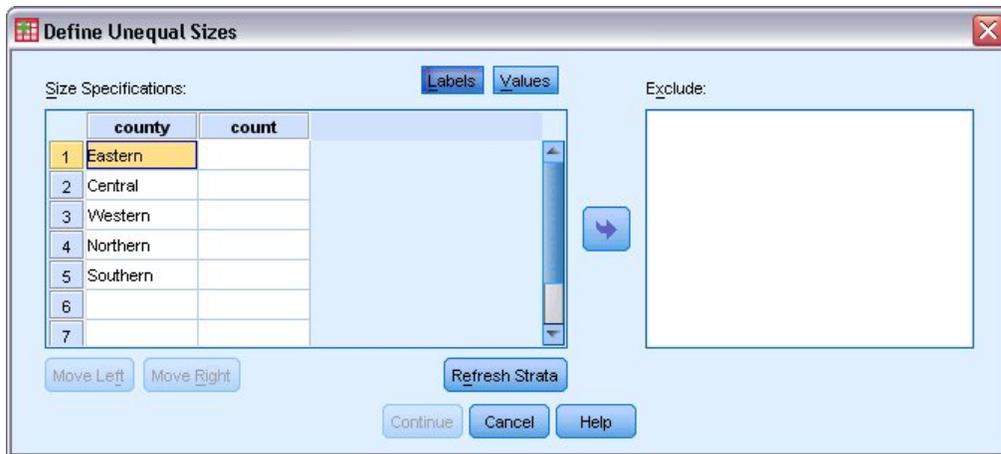


Figura 2. Caixa de diálogo Definir tamanhos desiguais

A caixa de diálogo Definir tamanhos desiguais permite inserir tamanhos por estrato.

Grade de especificações de tamanho. A grade exibe as classificações cruzadas de até cinco variáveis de estratos ou clusters — uma combinação de estratos/cluster por linha. As variáveis de grade elegíveis incluem todas as variáveis de estratificação dos estágios atuais e anteriores e todas as variáveis de cluster dos estágios anteriores. As variáveis podem ser reordenadas dentro da grade ou movidas para a lista de 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.

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

Assistente de preparação de análise: resumo do plano

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

Se não for possível incluir outro estágio, é provável que:

- Nenhuma variável de cluster foi especificada na etapa Variáveis de design.
- Você selecionou a estimativa WR na etapa Método de estimativa.
- Este é o terceiro estágio da análise, e o assistente suporta no máximo 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 em estágios no arquivo de plano existente, é possível salvar o plano editado em um novo arquivo ou substituir o arquivo existente. Ao incluir estágios sem fazer mudanças nos estágios existentes, o Assistente substitui automaticamente 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. A partir dos 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 basicamente as mesmas de um novo design. Para obter mais informações, consulte a Ajuda para etapas individuais.

5. Navegue até a etapa Concluir e especifique um novo nome para o arquivo de plano editado ou opte por substituir o arquivo de plano existente.

Opcionalmente, é possível remover estágios do plano.

Assistente de preparação de análise: resumo do plano

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

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

Plano de amostras complexas

Os procedimentos de análise de Amostras Complexas requerem especificações de análise de uma análise ou arquivo de plano de amostra para fornecer resultados válidos.

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

Probabilidades conjuntas. Para usar a estimativa WOR desigual para clusters desenhados usando um método PPS WOR, você precisa especificar um arquivo separado ou um conjunto de dados aberto contendo as probabilidades conjuntas. Este arquivo ou conjunto de dados é criado pelo assistente de amostragem durante a amostragem.

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, definidas por uma ou mais variáveis categóricas.

Exemplo. Usando o procedimento Frequências de Amostras Complexas, é possível obter estatísticas tabulares univariadas para o uso de vitaminas entre cidadãos dos EUA, com base nos resultados do 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 de população de células e porcentagens de tabela, além de erros padrão, intervalos de confiança, coeficientes de variação, efeitos de projeto, raízes quadradas de efeitos de projeto, valores cumulativos e contagens não ponderadas para cada

estimativa. Além disso, estatísticas de qui-quadrado e razão de verossimilhança são computadas para o teste de proporções celulares iguais.

Considerações de dados de Frequências de amostras complexas

Dados. As variáveis para as quais são produzidas tabelas de frequência devem ser categóricas. As variáveis de subpopulação podem ser 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 Frequências de amostras complexas

1. A partir dos menus, escolha:

Analisar > Amostras complexas > Frequências...

2. Selecione um arquivo de plano. Opcionalmente, selecione um arquivo de probabilidade conjuntas customizado.

3. Clique **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 de população de células e porcentagens da 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, usando o nível especificado.
- **Coefficiente de variação.** A razão entre o erro padrão da estimativa e a estimativa.
- **Contagem não ponderada.** O número de unidades usadas para calcular a estimativa.
- **Efeito do design.** A razão entre a variância da estimativa e a variância obtida assumindo que a amostra é uma amostra aleatória simples. Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.
- **Valores acumulativos.** A estimativa cumulativa por meio de cada valor da variável.

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

Valores omissos de amostras complexas

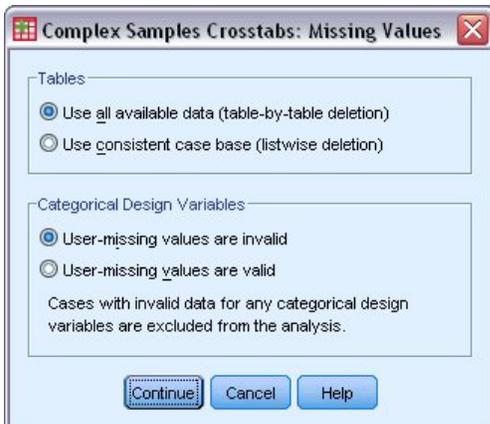


Figura 3. Caixa de diálogo Valores Omissos

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. Assim, os casos usados para calcular as estatísticas podem variar entre as tabelas de frequência ou tabulação cruzada.
- **Usar base de caso consistente.** Os valores ausentes são determinados em todas as variáveis. Assim, os casos usados para calcular as estatísticas são consistentes nas tabelas.

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

Opções de amostras complexas

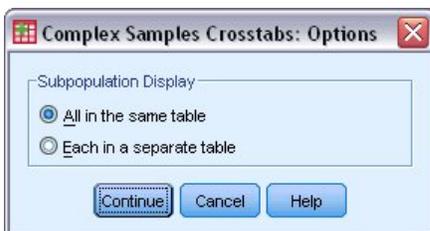


Figura 4. Caixa de diálogo Opções

Exibição de subpopulação. É possível optar por exibir subpopulações na mesma tabela ou em tabelas separadas.

Descrições de amostras complexas

O procedimento Descritivos de Amostras Complexas exibe estatísticas básicas univariadas para várias variáveis. Opcionalmente, é possível solicitar estatísticas por subgrupos, definidas por uma ou mais variáveis categóricas.

Exemplo. Usando o procedimento Descritivos de Amostras Complexas, é possível obter estatísticas descritivas univariadas para os níveis de atividade dos cidadãos dos EUA, com base nos resultados do National Health Interview Survey (NHIS) e com um plano de análise apropriado para esses 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 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. A partir dos menus, escolha:

Analisar > Amostras complexas > Descritivas...

2. Selecione um arquivo de plano. Opcionalmente, selecione um arquivo de probabilidades conjunta customizado.

3. Clique **Continuar**.

4. Na caixa de diálogo Descritivos de Amostras Complexas, 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 e somas 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, usando o nível especificado.
- **Coefficiente de variação.** A razão entre o erro padrão da estimativa e 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 entre a variância da estimativa e a variância obtida assumindo que a amostra é uma amostra aleatória simples. Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Esta é uma medida do efeito da especificação de um design complexo, em que valores além 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 ausentes são determinados variável por variável, portanto, os casos usados para calcular as estatísticas podem variar entre as variáveis de medida.
- **Assegure a base de caso consistente.** Os valores ausentes são determinados em todas as variáveis, portanto, os casos usados para calcular as estatísticas são consistentes.

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

Opções de amostras complexas

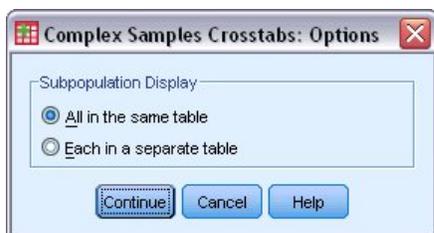


Figura 5. Caixa de diálogo Opções

Exibição de subpopulação. É possível optar por exibir subpopulações na mesma tabela ou em tabelas separadas.

Tabulações cruzadas de amostras complexas

O procedimento Tabelas cruzadas de amostras complexas produz tabelas de tabulação cruzada para pares de variáveis selecionadas e exibe estatísticas bidirecionais. Opcionalmente, é possível solicitar estatísticas por subgrupos, definidas por uma ou mais variáveis categóricas.

Exemplo. Usando o procedimento Tabelas cruzadas de amostras complexas, é possível obter estatísticas de classificação cruzada para frequência de tabagismo por uso de vitaminas de cidadãos dos EUA, com base nos resultados do 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 de preenchimento de células e porcentagens de linhas, colunas e tabelas, além de erros padrão, intervalos de confiança, coeficientes de variação, valores esperados, efeitos de projeto, raízes quadradas de efeitos de projeto, resíduos, resíduos ajustados e contagens não ponderadas para cada estimativa. A razão de chances, risco relativo e diferença de risco são calculados para tabelas 2 por 2. Além disso, as estatísticas de Pearson e razão de verossimilhança são computadas para o teste de independência das variáveis de linha e 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 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. A partir dos menus, escolha:

Analisar > Amostras complexas > Tabulações cruzadas...

2. Selecione um arquivo de plano. Opcionalmente, selecione um arquivo de probabilidades conjunta customizado.

3. Clique **Continuar**.

4. Na caixa de diálogo Tabelas cruzadas de amostras complexas, selecione pelo menos uma variável de linha e uma variável de 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. Esse grupo permite solicitar estimativas do tamanho da população de células e porcentagens de linhas, colunas e tabelas.

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

- **Erro padrão.** O erro padrão da estimativa.
- **Intervalo de confiança.** Um intervalo de confiança para a estimativa, usando o nível especificado.
- **Coefficiente de variação.** A razão entre o erro padrão da estimativa e a estimativa.
- **Valores esperados.** O valor esperado da estimativa, na hipótese de independência da variável linha e coluna.
- **Contagem não ponderada.** O número de unidades usadas para calcular a estimativa.
- **Efeito do design.** A razão entre a variância da estimativa e a variância obtida assumindo que a amostra é uma amostra aleatória simples. Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.
- **Residuais.** O valor esperado é o número de casos que você esperaria na célula se não houvesse relação entre as duas variáveis. Um resíduo positivo indica que há mais casos na célula do que haveria se as variáveis de linha e coluna fossem independentes.
- **Resíduos ajustados.** O resíduo de uma célula (observado menos o valor esperado) dividido por uma estimativa de 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 as variáveis de linha e coluna têm duas categorias cada. Cada um é uma medida da força 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 estimativa de risco relativo quando a ocorrência do fator é rara.
- **Risco relativo.** A razão entre o risco de um evento na presença do fator e 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 razão de verossimilhança da hipótese de que uma variável de linha e coluna são independentes. Testes separados são realizados para cada par de variáveis.

Valores omissos de amostras complexas

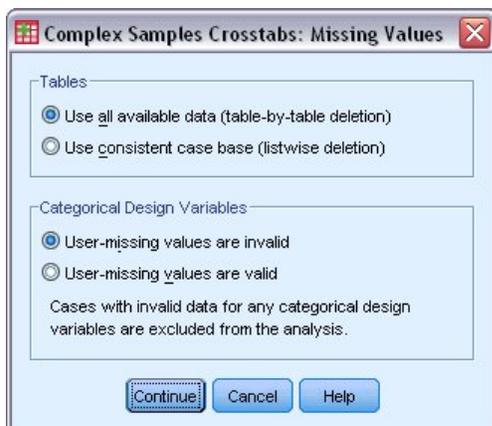


Figura 6. Caixa de diálogo Valores Omissos

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. Assim, os casos usados para calcular as estatísticas podem variar entre as tabelas de frequência ou tabulação cruzada.
- **Usar base de caso consistente.** Os valores ausentes são determinados em todas as variáveis. Assim, os casos usados para calcular as estatísticas são consistentes nas tabelas.

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

Opções de amostras complexas

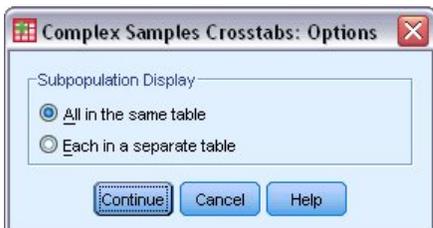


Figura 7. Caixa de diálogo Opções

Exibição de subpopulação. É possível optar por exibir subpopulações na mesma tabela ou em tabelas separadas.

Razões de amostras complexas

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

Exemplo. Usando o procedimento Índices de Amostras Complexas, é possível obter estatísticas descritivas para a razão entre o valor da propriedade atual e o último valor avaliado, com base nos resultados de uma pesquisa de opinião estadual realizada de acordo com um projeto complexo e com um plano de análise adequado 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. Os numeradores e denominadores devem ser variáveis de escala de valor positivo. As variáveis de subpopulação podem ser 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. A partir dos menus, escolha:

Analisar > Amostras complexas > Razões...

2. Selecione um arquivo de plano. Opcionalmente, selecione um arquivo de probabilidades conjunta customizado.

3. Clique **Continuar**.

4. Na caixa de diálogo Razões de amostras complexas, selecione pelo menos uma variável de numerador e uma variável de 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 da razão.

- **Erro padrão.** O erro padrão da estimativa.
- **Intervalo de confiança.** Um intervalo de confiança para a estimativa, usando o nível especificado.
- **Coefficiente de variação.** A razão entre o erro padrão da estimativa e 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 entre a variância da estimativa e a variância obtida assumindo que a amostra é uma amostra aleatória simples. Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Esta é uma medida do efeito da especificação de um design complexo, em que valores além 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 as estatísticas podem variar entre os pares numerador-denominador.
- **Assegure a base de caso consistente.** Os valores ausentes 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

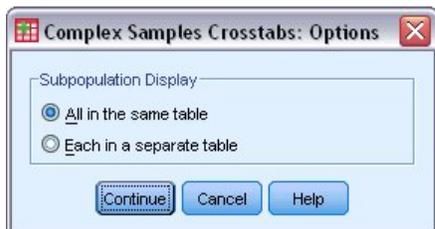


Figura 8. Caixa de diálogo Opções

Exibição de subpopulação. É possível optar por exibir subpopulações na mesma tabela ou em tabelas separadas.

Modelo linear geral de amostras complexas

O procedimento Modelo Linear Geral de Amostras Complexas (CSGLM) realiza análise de regressão linear, bem como análise de variância e covariância, para amostras extraídas por métodos de amostragem complexos. Opcionalmente, é possível solicitar análises para uma subpopulação.

Exemplo. Uma cadeia de supermercados pesquisou 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 que os clientes compram está relacionada ao valor que eles gastam em um mês, controlando o gênero do cliente e incorporando o design 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. Medidas de ajuste do modelo e estatísticas descritivas para as variáveis dependentes e independentes também estão disponíveis. Além disso, é possível solicitar médias marginais estimadas para níveis de fatores de modelo e interações de 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 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. A partir dos menus, escolha:

Analisar > Amostras complexas > Modelo linear geral...

2. Selecione um arquivo de plano. Opcionalmente, selecione um arquivo de probabilidades conjunta customizado.

3. Clique **Continuar**.

4. Na caixa de diálogo Modelo linear geral de amostras complexas, selecione uma variável dependente.

Opcionalmente, é possível:

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

Modelo linear geral de amostras complexas

Especificar Efeitos do Modelo. Por padrão, o procedimento cria um modelo de efeitos principais usando os fatores e covariáveis especificados na caixa de diálogo principal. Como alternativa, é possível criar um modelo personalizado que inclua 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 bidirecionais possíveis das variáveis selecionadas.

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

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

Todas 5 fatores. Cria todas as possíveis interações de cinco vias 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 do outro fator. Por exemplo, uma rede de supermercados pode seguir os hábitos de gastos de seus clientes em vários locais de loja. Como cada cliente frequenta somente um desses locais, o efeito do *Cliente* pode ser considerado **aninhado dentro** do efeito do *Local de armazenamento*.

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

Limitações. Os termos aninhados possuem as restrições a seguir:

- Todos os fatores em uma interação devem ser exclusivos. Dessa forma, se A for um fator, então, especificar A*A é inválido.

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

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 do 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 entre a variância da estimativa e a variância obtida assumindo que a amostra é uma amostra aleatória simples. Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Esta é uma medida do efeito da especificação de um design complexo, em que valores além 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 resumidas sobre a variável dependente, covariáveis e fatores.

Informações do plano de amostra. Exibe informações resumidas 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 as 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 desenho amostral, o valor é a diferença entre o número de unidades primárias de amostragem e o número de estratos na primeira etapa de amostragem. Como alternativa, é possível configurar graus de liberdade customizados especificando um número inteiro positivo.

Ajuste para Várias 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 que você escolha o método de ajuste.

- **Diferença menos significativa.** Este método não controla a probabilidade global de rejeitar as hipóteses de que alguns contrastes lineares são diferentes dos valores da hipótese nula.
- *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 pelo modelo para níveis de fatores e interações de fatores especificadas na subcaixa de diálogo Modelo. Também é possível solicitar que a média geral da população seja exibida.

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

Contraste. O contraste determina como os testes de hipóteses 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 de fator com o qual os outros são comparados.

Salvar modelo linear geral de amostras complexas

Salvar variáveis. Este grupo permite que você salve os valores preditos do modelo e os resíduos 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 de matriz é mostrada a seguir.

- **rowtype_.** Recebe valores (e rótulos de valor), COV (Covariâncias), CORR (Correlações), EST (Estimativas de parâmetros), SE (Erros padrão), SIG (Níveis de significância) e DF (Graus de liberdade do projeto de amostragem). Há um caso separado com o tipo de linha COV (ou CORR) para cada parâmetro do modelo, além de um caso separado para cada um dos outros tipos de linha.
- **varname_.** Assume os valores P1, P2, ..., correspondentes a uma lista ordenada de todos os parâmetros do modelo, para os tipos de linha COV ou CORR, com rótulos de valor correspondentes às cadeias de parâmetros mostradas na tabela de estimativas de parâmetros. As células estão em branco para outros tipos de linha.
- **P1, P2, ...** Essas variáveis correspondem a uma lista ordenada de todos os parâmetros do modelo, com rótulos de variáveis correspondentes às cadeias de parâmetros mostradas na tabela de estimativas de parâmetros, e assumem valores de acordo com o tipo de linha. Para parâmetros redundantes, todas as covariâncias são configuradas como zero; as correlações são configuradas para o valor omisso do

sistema; todas as estimativas de parâmetros são configuradas em zero; e todos os erros padrão, níveis de significância e graus de liberdade residuais são configuradas para 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. Para obter mais informações.

Opções de modelo linear geral de amostras complexas

Valores Omissos de Usuário. Todas as variáveis de projeto, assim 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 que você decida se os valores ausentes do usuário são tratados como válidos entre as variáveis de estratos, cluster, subpopulação e de fator.

Intervalo de Confiança. Este é o nível de intervalo de confiança para estimativas de coeficientes 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).
- Fixar 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).
- Produza uma tabela de funções estimáveis gerais (usando o subcomando PRINT).

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

Regressão logística de amostras complexas

O procedimento Regressão Logística de Amostras Complexas executa a análise de regressão logística em uma variável dependente binária ou multinomial para amostras extraídas por métodos de amostragem complexos. Opcionalmente, é possível solicitar análises para uma subpopulação.

Exemplo. Um responsável pelo empréstimo coletou registros anteriores de clientes que concederam empréstimos em várias agências diferentes, de acordo com um design complexo.que fizeram empréstimos em várias agências diferentes, de acordo com um design complexo. Ao incorporar o desenho da amostra, o oficial quer ver se a probabilidade de inadimplência de um cliente está relacionada à idade, histórico de emprego e valor da dívida de crédito.

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 sobre 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 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 Amostras Complexas Regressão Logística

1. A partir dos menus, escolha:

Analisar > Amostras complexas > Regressão logística...

2. Selecione um arquivo de plano. Opcionalmente, selecione um arquivo de probabilidades conjunta customizado.

3. Clique **Continuar**.

4. Na caixa de diálogo Regressão logística de amostras complexas, selecione uma variável dependente.

Opcionalmente, é possível:

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

Categoria de referência de regressão logística de amostras complexas

Por padrão, o procedimento de Regressão Logística de Amostras Complexas torna a categoria de maior valor a categoria de referência. Essa caixa de diálogo permite especificar o valor mais alto, o valor mais baixo ou uma categoria personalizada como a categoria de referência.

Modelo de regressão logística de Amostras Complexas

Especificar Efeitos do Modelo. Por padrão, o procedimento cria um modelo de efeitos principais usando os fatores e covariáveis especificados na caixa de diálogo principal. Como alternativa, é possível criar um modelo personalizado que inclua 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 bidirecionais possíveis das variáveis selecionadas.

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

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

Todas 5 fatores. Cria todas as possíveis interações de cinco vias 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 do outro fator. Por exemplo, uma rede de supermercados pode seguir os hábitos de gastos de seus clientes em vários locais de loja. Como cada cliente frequenta somente um desses locais, o efeito do *Cliente* pode ser considerado **aninhado dentro** do efeito do *Local de armazenamento*.

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

Limitações. Os termos aninhados possuem as restrições a seguir:

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

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 prevista pelo modelo na variável dependente.

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 tenha boas propriedades para testes estatísticos, a estimativa e exponencial, 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 entre a variância da estimativa e a variância obtida assumindo que a amostra é uma amostra aleatória simples. Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.

estatísticas básicas para variáveis de modelo. Exibe informações resumidas sobre a variável dependente, covariáveis e fatores.

Informações do plano de amostra. Exibe informações resumidas 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 as 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 desenho amostral, o valor é a diferença entre o número de unidades primárias de amostragem e o número de

estratos na primeira etapa de amostragem. Como alternativa, é possível configurar graus de liberdade customizados especificando um número inteiro positivo.

Ajuste para Várias 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 que você escolha o método de ajuste.

- **Diferença menos significativa.** Este método não controla a probabilidade global de rejeitar as hipóteses de que alguns contrastes lineares são diferentes dos valores da hipótese nula.
- *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 as 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 de alteração especificadas para as chances na média.

Ao calcular as razões de chances para um fator ou covariável, o procedimento fixa todos os outros fatores em seus níveis mais altos e todas as outras covariáveis em suas médias. Se um fator ou covariável interage com outros preditores no modelo, as razões de chances dependem não apenas da mudança na variável especificada, mas também dos valores das variáveis com as quais 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 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 de matriz é mostrada a seguir.

- **rowtype_.** Recebe valores (e rótulos de valor), COV (Covariâncias), CORR (Correlações), EST (Estimativas de parâmetros), SE (Erros padrão), SIG (Níveis de significância) e DF (Graus de liberdade do projeto de amostragem). Há um caso separado com o tipo de linha COV (ou CORR) para cada parâmetro do modelo, além de um caso separado para cada um dos outros tipos de linha.
- **varname_.** Assume os valores P1, P2, ..., correspondentes a uma lista ordenada de todos os parâmetros do modelo, para os tipos de linha COV ou CORR, com rótulos de valor correspondentes às cadeias de parâmetros mostradas na tabela de estimativas de parâmetros. As células estão em branco para outros tipos de linha.
- **P1, P2, ...** Essas variáveis correspondem a uma lista ordenada de todos os parâmetros do modelo, com rótulos de variáveis correspondentes às cadeias de parâmetros mostradas na tabela de estimativas de parâmetros, e assumem valores de acordo com o tipo de linha. Para parâmetros redundantes, todas

as covariâncias são configuradas como zero; as correlações são configuradas para o valor omisso do sistema; todas as estimativas de parâmetros são configuradas em zero; e todos os erros padrão, níveis de significância e graus de liberdade residuais são configuradas para 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. para obter mais informações.

Opções de Regressão logística de amostras complexas

Estimação. Este grupo fornece a você o controle de vários critérios usados na estimativa do 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.** A cada iteração, o tamanho da etapa é reduzido por um fator de 0,5 até que a probabilidade de log aumente ou a etapa máxima seja atingida. Especifique um número inteiro positivo.
- **Limitar iterações com base na mudança nas estimativas paramétrica.** Quando selecionado, o algoritmo para após uma iteração na qual a mudança absoluta ou relativa nas estimativas de parâmetro é menor que o valor especificado, que deve ser não negativo.
- **Limitar iterações com base na mudança no log da verossimilhança.** Quando selecionado, o algoritmo para após uma iteração na qual a alteração absoluta ou relativa na função de probabilidade de log 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 classifique corretamente todos os casos.
- **Exibir histórico de iteração.** Displays parameter estimates and statistics at every n iterations beginning with the 0th iteration (the initial estimates). 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 projeto, assim 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 que você decida se os valores ausentes do usuário são tratados como válidos entre as variáveis de estratos, cluster, subpopulação e de fator.

Intervalo de Confiança. Este é o nível do intervalo de confiança para estimativas de coeficientes, estimativas de coeficientes exponenciados 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).
- Produza uma tabela de funções estimáveis gerais (usando o subcomando PRINT).

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

Regressão ordinal de amostras complexas

O procedimento Regressão Ordinal de Amostras Complexas executa a análise de regressão em uma variável dependente binária ou ordinal para amostras extraídas por métodos de amostragem complexos. Opcionalmente, é possível solicitar análises para uma subpopulação.

Exemplo. Os deputados que consideram um projeto de lei antes da legislatura estão interessados em saber se há apoio público ao projeto e como o apoio ao projeto está relacionado à demografia dos eleitores. Pesquisadores projetam e conduzem entrevistas de acordo com um projeto de amostragem complexo. Usando Regressão Ordinal de Amostras Complexas, é possível ajustar um modelo para o nível de suporte ao projeto de lei com base na demografia do eleitor.

Considerações sobre 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 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 regressão ordinal de amostras complexas

1. A partir dos menus, escolha:

Analisar > Amostras complexas > Regressão ordinal...

2. Selecione um arquivo de plano. Opcionalmente, selecione um arquivo de probabilidades conjunta customizado.

3. Clique **Continuar**.

4. Na caixa de diálogo Regressão Ordinal de Amostras Complexas, selecione uma variável dependente.

Opcionalmente, é possível:

- Selecione variáveis para fatores e covariáveis, conforme apropriado para seus dados.
- Especifique uma variável para definir uma subpopulação. A análise é realizada apenas para a categoria selecionada da variável de subpopulação, embora as variâncias ainda sejam adequadamente estimadas com base em todo o conjunto de dados.
- Selecione uma função de ligação.

Função de ligação. A função de ligação é uma transformação das probabilidades cumulativas que permite estimar o modelo. As cinco funções de link a seguir estão disponíveis.

- **Logit.** $f(x) = \log(x/(1-x))$. Normalmente usada para categorias distribuídas uniformemente.
- **Log-log complementar.** $f(x) = \log(-\log(1-x))$. Normalmente usada quando as categorias mais altas são mais prováveis.
- **Log-log negativo.** $f(x) = -\log(-\log(x))$. Normalmente usada quando as categorias mais baixas são mais prováveis.
- **Probit.** $f(x) = \delta^{-1}(x)$. Normalmente usada quando a variável latente é normalmente distribuída.
- **Cauchit (inverse Cauchy).** $f(x) = \tan(\pi(x-0.5))$. Normalmente usada quando a variável latente tem muitos valores extremos.

Probabilidades de resposta de regressão ordinal de amostras complexas

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

Modelo de regressão ordinal de amostras complexas

Especificar Efeitos do Modelo. Por padrão, o procedimento cria um modelo de efeitos principais usando os fatores e covariáveis especificados na caixa de diálogo principal. Como alternativa, é possível criar um modelo personalizado que inclua 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 bidirecionais possíveis das variáveis selecionadas.

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

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

Todas 5 fatores. Cria todas as possíveis interações de cinco vias 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 do outro fator. Por exemplo, uma rede de supermercados pode seguir os hábitos de gastos de seus clientes em vários locais de loja. Como cada cliente frequenta somente um desses locais, o efeito do *Cliente* pode ser considerado **aninhado dentro** do efeito do *Local de armazenamento*.

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

Limitações. Os termos aninhados possuem as restrições a seguir:

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

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 prevista pelo modelo na variável dependente.

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 tenha boas propriedades para testes estatísticos, a estimativa e exponencial, 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 entre a variância da estimativa e a variância obtida assumindo que a amostra é uma amostra aleatória simples. Esta é uma medida do efeito da especificação de um design complexo, em que valores além 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, onde valores mais distantes de 1 indicam maiores efeitos.

Linhas paralelas. Este grupo permite solicitar estatísticas associadas a um modelo com linhas não paralelas onde 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 cumulativas. O modelo com linhas não paralelas é estimado e o teste de Wald de parâmetros iguais é aplicado.
- **Estimativas paramétrica.** Exibe estimativas dos coeficientes e erros padrão para o modelo com linhas não paralelas.
- **Covariâncias de estimativas paramétrica.** Exibe uma estimativa da matriz de covariância para os coeficientes do modelo com linhas não paralelas.

estatísticas básicas para variáveis de modelo. Exibe informações resumidas sobre a variável dependente, covariáveis e fatores.

Informações do plano de amostra. Exibe informações resumidas 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 as 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 desenho amostral, o valor é a diferença entre o número de unidades primárias de amostragem e o número de estratos na primeira etapa de amostragem. Como alternativa, é possível configurar graus de liberdade customizados especificando um número inteiro positivo.

Ajuste para Várias 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 que você escolha o método de ajuste.

- **Diferença menos significativa.** Este método não controla a probabilidade global de rejeitar as hipóteses de que alguns contrastes lineares são diferentes dos valores da hipótese nula.
- *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 as razões de chances cumulativas estimadas pelo modelo para fatores e covariáveis especificados. Este recurso está disponível apenas para modelos que usam a função de ligação Logit. Uma única razão de chances cumulativas é calculada para todas as categorias da variável dependente, exceto a última; o modelo de probabilidades proporcionais postula que todas são iguais.

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

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

Ao calcular as razões de chances para um fator ou covariável, o procedimento fixa todos os outros fatores em seus níveis mais altos e todas as outras covariáveis em suas médias. Se um fator ou covariável interage com outros preditores no modelo, as razões de chances dependem não apenas da mudança na variável especificada, mas também dos valores das variáveis com as quais 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.

Salvar Regressão Ordinal de Amostras Complexas

Salvar variáveis. Este grupo permite que você salve a categoria predita pelo modelo, a probabilidade da categoria predita, a probabilidade da categoria observada, as probabilidades cumulativas e as probabilidades previstas como novas variáveis no conjunto de dados ativo.

Exportar modelo como ddos 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 de matriz é mostrada a seguir.

- **rowtype_.** Recebe valores (e rótulos de valor), COV (Covariâncias), CORR (Correlações), EST (Estimativas de parâmetros), SE (Erros padrão), SIG (Níveis de significância) e DF (Graus de liberdade do projeto de amostragem). Há um caso separado com o tipo de linha COV (ou CORR) para cada parâmetro do modelo, além de um caso separado para cada um dos outros tipos de linha.
- **varname_.** Assume os valores P1, P2, ..., correspondentes a uma lista ordenada de todos os parâmetros do modelo, para os tipos de linha COV ou CORR, com rótulos de valor correspondentes às cadeias de parâmetros mostradas na tabela de estimativas de parâmetros. As células estão em branco para outros tipos de linha.
- **P1, P2, ...** Essas variáveis correspondem a uma lista ordenada de todos os parâmetros do modelo, com rótulos de variáveis correspondentes às cadeias de parâmetros mostradas na tabela de estimativas de parâmetros, e assumem valores de acordo com o tipo de linha. Para parâmetros redundantes, todas as covariâncias são configuradas como zero; as correlações são configuradas para o valor omisso do sistema; todas as estimativas de parâmetros são configuradas em zero; e todos os erros padrão, níveis de significância e graus de liberdade residuais são configuradas para 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. para obter mais informações.

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 de escoragem de Fisher são executadas antes de alternar para o método de Newton-Raphson. Se a convergência for

alcançada durante a fase de escoragem de Fisher do método híbrido antes que o número máximo de iterações de Fisher seja alcançado, o algoritmo continua com o método de Newton-Raphson.

Estimação. Este grupo fornece a você o controle de vários critérios usados na estimativa do 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.** A cada iteração, o tamanho da etapa é reduzido por um fator de 0,5 até que a probabilidade de log aumente ou a etapa máxima seja atingida. Especifique um número inteiro positivo.
- **Limitar iterações com base na mudança nas estimativas paramétrica.** Quando selecionado, o algoritmo para após uma iteração na qual a mudança absoluta ou relativa nas estimativas de parâmetro é menor que o valor especificado, que deve ser não negativo.
- **Limitar iterações com base na mudança no log da verossimilhança.** Quando selecionado, o algoritmo para após uma iteração na qual a alteração absoluta ou relativa na função de probabilidade de log 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 classifique corretamente todos os casos.
- **Exibir histórico de iteração.** Displays parameter estimates and statistics at every n iterations beginning with the 0th iteration (the initial estimates). 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 projeto de escala, assim 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 que você decida se os valores ausentes do usuário são tratados como válidos entre as variáveis de estratos, cluster, subpopulação e de fator.

Intervalo de Confiança. Este é o nível do intervalo de confiança para estimativas de coeficientes, estimativas de coeficientes exponenciados 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 ODDS RATIOS).
- Use valores não rotulados como categorias de referência customizada para fatores quando razões de chances são solicitadas (usando o subcomando ODDS RATIOS).
- Especifique um valor de tolerância para a verificação de singularidade (usando o subcomando CRITERIA).
- Produza uma tabela de funções estimáveis gerais (usando o subcomando PRINT).
- Salvar mais de 25 variáveis de probabilidade (usando o subcomando SAVE).

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

Regressão de Cox de amostras complexas

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

Exemplos. Uma agência governamental de cumprimento da lei está preocupada com as taxas de reincidência em sua área de jurisdição. Uma das medidas de reincidência é o tempo até a segunda prisão para os infratores. A agência gostaria de modelar o tempo para voltar a deter usando a Regressão de Cox,

mas está preocupada que a suposição de riscos proporcionais seja inválida em todas as categorias de idade.

Pesquisadores médicos estão investigando os tempos de sobrevivência para pacientes que saem de um programa de reabilitação pós-AVC isquêmico. Existe a possibilidade de múltiplos casos por sujeito, uma vez que as histórias dos pacientes mudam à medida que a ocorrência de eventos significativos não-morte são observadas e os tempos desses eventos são registrados. A amostra também é truncada à esquerda no sentido de que os tempos de sobrevivência observados são "inflados" pela duração da reabilitação, pois enquanto o início do risco começa no momento do AVC isquêmico, apenas os pacientes que sobrevivem após o programa de reabilitação estão em a amostra.

Considerações dos dados de regressão de Cox de amostras complexas

Tempo de sobrevivência. O procedimento aplica a regressão de Cox à análise dos tempos de sobrevivência - isto é, o período de tempo antes da ocorrência de um evento. Existem duas maneiras de especificar o tempo de sobrevivência, dependendo da hora de início do intervalo:

- **Tempo=0.** Normalmente, você terá informações completas sobre o início do intervalo para cada assunto e terá simplesmente uma variável contendo os horários de término (ou crie uma única variável com os horários de término das variáveis Data e 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 a sua amostra incluir apenas pacientes que sobreviveram ao programa de reabilitação, sua amostra será truncada à esquerda no sentido de que os tempos de sobrevivência observados são "inflados" pela duração da reabilitação. É possível explicar isso especificando a hora em que eles saíram da reabilitação como o horário de entrada no estudo.

Variáveis de Data e Hora. As variáveis de data e hora não podem ser usadas para definir diretamente o início e o fim do intervalo; se você tiver variáveis de data e hora, você deve usá-las para criar variáveis contendo tempos de sobrevivência. Se não houver truncamento à esquerda, basta criar uma variável contendo os tempos de término com base na diferença entre a data de entrada no estudo e a data de observação. Se houver truncamento à esquerda, crie uma variável contendo os horários de início, com base na diferença entre a data de início do estudo e a data de entrada, e uma variável contendo os horários de término, com base na diferença entre a data de início do estudo e a data da observação. para obter mais informações.

Status do evento. Você precisa de uma variável que registre se o sujeito experimentou o evento de interesse dentro do intervalo. Os sujeitos para os quais o evento não ocorreu são censurados à direita.

Identificador de sujeito. É possível incorporar facilmente preditores dependentes de tempo e constantes por partes dividindo as observações de um único sujeito em múltiplos casos. Por exemplo, se você estiver analisando os tempos de sobrevivência de pacientes pós-AVC, as variáveis que representam seu histórico médico devem ser úteis como preditores. Com o tempo, eles podem experimentar grandes eventos médicos que alteram 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 do tempo constantes por partes

ID do paciente	Horário de encerramento	Status	Histórico de ataque cardíaco anterior	Histórico de hemorragia anterior
1	5	Ataque cardíaco	Não	Não
1	7	Hemorragia	Sim	Não
1	8	Morto	Sim	Sim
2	24	Morto	Não	Não

Tabela 1. Estrutura de dados para incorporar preditores dependentes do tempo constantes por partes (continuação)

ID do paciente	Horário de encerramento	Status	Histórico de ataque cardíaco anterior	Histórico de hemorragia anterior
3	8	Ataque cardíaco	Não	Não
3	15	Morto	Sim	Não

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](#).

Normalmente, os modelos de regressão de Cox assumem riscos proporcionais, ou seja, a proporção de riscos de um caso para outro não deve variar ao longo do tempo. Se essa suposição não for válida, talvez seja necessário incluir preditores dependentes do tempo no modelo.

Análise de Kaplan-Meier. Se você não selecionar nenhum preditor (ou não inserir nenhum preditor selecionado no modelo) e escolher o método de limite do produto para calcular a curva de sobrevivência da linha de base na guia Opções, o procedimento executará uma análise de sobrevivência do tipo Kaplan-Meier.

Para obter regressão de Cox de amostras complexas

1. A partir dos menus, escolha:

Analisar > Amostras complexas > Regressão de Cox...

2. Selecione um arquivo de plano. Opcionalmente, selecione um arquivo de probabilidades conjunta customizado.

3. Clique **Continuar**.

4. Especifique o tempo de sobrevivência selecionando os horários de entrada e de 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 assunto identificador.

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 em uma lista de valores com rótulos de valor definidos.
- **Intervalo de valores.** Especifique um intervalo de valores inserindo os valores mínimo e máximo ou selecionando valores de uma lista com rótulos de valores definidos.

Preditores

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

Fatores. Os 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. Existem certas situações em que a suposição de riscos proporcionais não é válida. Ou seja, as razões de risco mudam ao longo do tempo; os valores de um (ou mais) de seus preditores são diferentes em diferentes momentos. Nesses casos, é necessário especificar preditores dependentes de tempo. Consulte o tópico [“Definir preditor dependente de tempo”](#) na página 34 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 depende da variável de tempo interna, $T_$. É possível usar essa variável para definir covariáveis dependentes do 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 funções mais complexas também podem ser especificadas.
- Algumas variáveis podem ter valores diferentes em diferentes períodos de tempo, mas não estão sistematicamente relacionadas ao 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. Usando uma série de expressões lógicas, é possível criar seu preditor dependente do tempo a partir de um conjunto de medições. Por exemplo, se você tiver a pressão arterial medida uma vez por semana durante as quatro semanas do seu estudo (identificada como $BP1$ a $BP4$), é possível definir seu preditor dependente de tempo como $(T_ < 1) * BP1 + (T_ \geq 1 \ \& \ T_ < 2) * BP2 + (T_ \geq 2 \ \& \ T_ < 3) * BP3 + (T_ \geq 3 \ \& \ T_ < 4) * BP4$. Observe que exatamente um dos termos entre parênteses será igual a 1 para qualquer caso e o restante será igual a 0. Em outras palavras, esta função significa que se o tempo for inferior a 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 [“Regressão de Cox de amostras complexas”](#) na página 31 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 pode inseri-la diretamente na área de texto Expressão Numérica. Observe que as constantes de cadeia de caracteres devem ser colocadas entre aspas ou apóstrofes e as constantes numéricas devem ser digitadas no formato americano, com o ponto como delimitador decimal. A variável resultante recebe o nome que você especifica e deve ser incluída como um fator ou covariável na guia Preditores.

Subgrupos

Camada de linha de base. Uma função de risco e sobrevivência de de linha de base separada é calculada para cada valor dessa variável, enquanto um único conjunto de coeficientes do modelo é estimado entre os estratos.

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

Modelo

Especificar Efeitos do Modelo. Por padrão, o procedimento cria um modelo de efeitos principais usando os fatores e covariáveis especificados na caixa de diálogo principal. Como alternativa, é possível criar um modelo personalizado que inclua 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 bidirecionais possíveis das variáveis selecionadas.

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

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

Todas 5 fatores. Cria todas as possíveis interações de cinco vias 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 do outro fator. Por exemplo, uma rede de supermercados pode seguir os hábitos de gastos de seus clientes em vários locais de loja. Como cada cliente frequenta somente um desses locais, o efeito do *Cliente* pode ser considerado **aninhado dentro** do efeito do *Local de armazenamento*.

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

Limitações. Os termos aninhados possuem as restrições a seguir:

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

Estatísticas

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

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

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

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 tenha boas propriedades para testes estatísticos, a estimativa e exponencial, 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 entre a variância da estimativa e a variância obtida assumindo que a amostra é uma amostra aleatória simples. Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.
- **Raiz quadrada de efeito do design.** Esta é uma medida do efeito da especificação de um design complexo, em que valores além de 1 indicam efeitos maiores.

Suposições de modelo. Este grupo permite que você produza 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 **identity**, $TF=T_$. 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 **rank**, TF é a ordem de classificação de $T_$ entre os tempos finais observados.
- **Estimativas paramétrica para o modelo alternativo.** Exibe a estimativa, o erro padrão e o intervalo de confiança para cada parâmetro no modelo alternativo.
- **Matriz de covariâncias para o modelo alternativo.** Exibe a matriz de covariâncias estimadas entre os 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 cumulativo da linha de base junto 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 Plots permite solicitar gráficos da função de risco, função de sobrevivência, log-minus-log da função de sobrevivência e um menos a função de sobrevivência. Também é possível optar por plotar intervalos de confiança ao longo das funções especificadas; o nível de confiança é definido na guia Opções.

Padrões de preditor. Também é possível especificar um padrão de valores preditores a serem usados para os gráficos solicitados e o arquivo de sobrevivência exportado na guia Exportar. Observe que essas opções não estarão disponíveis se os preditores dependentes de tempo definidos na guia Preditores estiverem incluídos no modelo.

- **Plotar fatores em.** Por padrão, cada fator é avaliado em seu nível mais alto. Insira ou selecione um nível diferente, se desejar. Como alternativa, é possível optar por traçar linhas separadas para cada nível de um único fator marcando a caixa de seleção desse fator.
- **Plotar covariáveis em.** Cada covariável é avaliada por sua média. Insira ou selecione um valor diferente, se desejar.

Testes de hipótese

Estatística do teste. Este grupo permite selecionar o tipo de estatística usada para testar as 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 desenho amostral, o valor é a diferença entre o número de unidades primárias de amostragem e o número de estratos na primeira etapa de amostragem. Como alternativa, é possível configurar graus de liberdade customizados especificando um número inteiro positivo.

Ajuste para Várias 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 que você escolha o método de ajuste.

- **Diferença menos significativa.** Este método não controla a probabilidade global de rejeitar as hipóteses de que alguns contrastes lineares são diferentes dos valores da hipótese nula.
- *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

Salvar variáveis. Este grupo permite que você salve variáveis relacionadas ao modelo no conjunto de dados ativo para uso posterior em diagnósticos e relatórios de resultados. Observe que nenhum deles está disponível quando os preditores dependentes do tempo são incluídos no modelo.

- **Função de sobrevivência.** Salva a probabilidade de sobrevivência (o valor da função de sobrevivência) no tempo 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 tempo 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 tempo observado e os valores do preditor para cada caso.
- **Função de risco acumulativo.** Salva o risco cumulativo, ou $-\ln(\text{sobrevivência})$, no tempo observado e valores preditores 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 cumulativo no tempo 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 cumulativo no tempo 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 linha de referência. No modelo de riscos proporcionais, esse valor é constante ao longo do tempo.
- **Resíduo de Schoenfeld.** Para cada caso não censurado e cada parâmetro não redundante no modelo, o resíduo de Schoenfeld é a diferença entre o valor observado do preditor associado ao parâmetro do modelo e o valor esperado do preditor para casos no risco definido no horário do evento observado. 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 computados apenas para casos não censurados.
- **Resíduo de Martingale.** Para cada caso, o resíduo martingale é a diferença entre a censura observada (0 se censurada, 1 se não) e a expectativa de um evento durante o tempo de observação.
- **Resíduo de deviance.** Resíduos de desvio são resíduos de martingale "ajustados" para parecerem mais simétricos em torno de 0. Gráficos de resíduos de desvio contra preditores não devem revelar padrões.
- **Resíduo de Cox-Snell.** Para cada caso, o 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 score.** Para cada caso e cada parâmetro não redundante no modelo, o score residual é a contribuição do caso para a primeira derivada 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 DFBeta aproxima a mudança no valor da estimativa do parâmetro quando o caso é removido do modelo. Casos com resíduos DFBeta relativamente grandes podem estar exercendo influência indevida na análise. Uma variável separada é salva para cada parâmetro não redundante no modelo.
- **Resíduos agregados.** Quando vários casos representam um único sujeito, o resíduo agregado para um sujeito é simplesmente a soma dos resíduos de caso correspondentes sobre todos os casos pertencentes ao mesmo sujeito. Para o resíduo de Schoenfeld, a versão agregada é a mesma da versão não agregada, pois o resíduo de Schoenfeld é definido apenas para casos não censurados. Esses resíduos só estão disponíveis quando um assunto identificador é especificado na guia Hora e Evento.

Nomes de Variáveis Salvas. A geração automática de nomes garante que você mantenha todo o seu trabalho. Os 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 de matriz é mostrada a seguir.

- **rowtype_.** Recebe valores (e rótulos de valor), COV (Covariâncias), CORR (Correlações), EST (Estimativas de parâmetros), SE (Erros padrão), SIG (Níveis de significância) e DF (Graus de liberdade do projeto de amostragem). Há um caso separado com o tipo de linha COV (ou CORR) para cada parâmetro do modelo, além de um caso separado para cada um dos outros tipos de linha.
- **varname_.** Assume os valores P1, P2, ..., correspondentes a uma lista ordenada de todos os parâmetros do modelo, para os tipos de linha COV ou CORR, com rótulos de valor correspondentes às cadeias de parâmetros mostradas na tabela de estimativas de parâmetros. As células estão em branco para outros tipos de linha.
- **P1, P2, ...** Essas variáveis correspondem a uma lista ordenada de todos os parâmetros do modelo, com rótulos de variáveis correspondentes às cadeias de parâmetros mostradas na tabela de estimativas de parâmetros, e assumem valores de acordo com o tipo de linha. Para parâmetros redundantes, todas as covariâncias são configuradas como zero; as correlações são configuradas para o valor omissivo do sistema; todas as estimativas de parâmetros são configuradas em zero; e todos os erros padrão, níveis de significância e graus de liberdade residuais são configuradas para 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 de matriz é mostrada 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_#.#, ...** Função de sobrevivência avaliada em cada um dos padrões de previsão especificados na guia Gráficos e nos limites superior e inferior de seus intervalos de confiança. Consulte a tabela de valores de 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_#.#, ...** Função de risco cumulativo avaliada em cada um dos padrões de previsão especificados na guia Gráficos e nos limites superior e inferior de seus intervalos de confiança. Consulte a tabela de valores de 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. para obter mais informações.

Opções

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

- **Máximo de iterações.** O número máximo de iterações que o algoritmo executará. Especifique um número inteiro não negativo.
- **Divisão máxima do passo pela metade.** A cada iteração, o tamanho da etapa é reduzido por um fator de 0,5 até que a probabilidade de log aumente ou a etapa máxima seja atingida. Especifique um número inteiro positivo.
- **Limitar iterações com base na mudança nas estimativas paramétrica.** Quando selecionada, o algoritmo é interrompido após uma iteração na qual uma mudança absoluta ou relativa nas estimativas de parâmetro for menor que o valor especificado, que deve ser positivo.
- **Limitar iterações com base na mudança no 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 de parâmetro e verossimilhança de pseudo log da e imprime a última avaliação da alteração nas estimativas de parâmetro e verossimilhança de pseudo log. 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 há tempos de falha observados empatados, um desses métodos é usado para quebrar as camadas. O método Efron é mais caro computacionalmente.

Funções de sobrevivência. Esses controles especificam critérios para cálculos envolvendo 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 logarítmica ou uma transformação log-menos-log. Apenas a transformação log-menos-log garante que os limites do intervalo de confiança ficarão entre 0 e 1, mas a transformação logarítmica geralmente parece ter um desempenho "melhor".

Valores omissos do usuário. Todas as variáveis devem ter valores válidos para que um caso seja incluído na análise. Esses controles permitem que você decida se os valores omissos do usuário são tratados como válidos entre modelos categóricos (incluindo fatores, eventos, estratos e variáveis de subpopulação) e variáveis de design de amostragem.

Intervalo de confiança(%). Este é o nível de intervalo de confiança usado para estimativas de coeficientes, estimativas de coeficientes exponenciados, estimativas de função de sobrevivência e estimativas de função de risco cumulativo. 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 /PRINT GEF).
- Padrões preditores múltiplos (usando subcomandos 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 de Sintaxe de Comando* para obter informações de sintaxe completa.

Avisos

Estas informações foram desenvolvidas para produtos e serviços oferecidos nos EUA. Esse material pode estar disponível a partir da IBM em outros idiomas. Entretanto, pode ser necessário que possua uma cópia do produto ou versão de produto nesse idioma a fim de acessá-lo.

É possível que a IBM não ofereça os produtos, serviços ou recursos discutidos nesta publicação em outros países. Consulte um representante IBM local para obter informações sobre produtos e serviços disponíveis atualmente em sua área. Qualquer referência a um produto, programa ou serviço IBM não está destinado a declarar ou implicar que apenas esse produto, programa ou serviço IBM possa ser usado. 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. É possível enviar consultas sobre licenças, por escrito, para:

Gerência de Relações Comerciais e Industriais da IBM Brasil

*Av. Pasteur, 138-146, Botafogo
Botafogo
Rio de Janeiro, RJCEP 22290-240*

Para consultas sobre 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 consultas sobre licença, por escrito, para:

Intellectual Property Licensing

*Legal and Intellectual Property Law
IBM Japan Ltd.*

19-21, Nihonbashi-Hakozakicho, Chuo-kuTokyo 103-8510, Japan

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

Essas informações podem conter imprecisões técnicas ou erros tipográficos. Periodicamente, são feitas mudanças nas informações aqui contidas; tais mudanças serão incorporadas em novas edições da publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar os produtos e/ou programas 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 desse produto IBM e a utilização desses websites é de inteira responsabilidade do Cliente.

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

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

*Av. Pasteur, 138-146, Botafogo
Botafogo
Rio de Janeiro, RJCEP 22290-240*

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

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

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

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

LICENÇA DE COPYRIGHT:

Estas informações contêm programas de aplicativos de amostra na linguagem fonte, ilustrando as técnicas de programação em diversas plataformas operacionais. O Cliente pode copiar, modificar e distribuir estes programas de exemplo 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 nenhum tipo. A IBM não será responsabilizada por quaisquer danos decorrentes do uso dos programas de amostra.

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

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

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

Marcas comerciais

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

Adobe, o logotipo Adobe, PostScript e o logotipo PostScript são marcas ou marcas registradas 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 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 de suas afiliadas.

Índice remissivo

A

amostragem
 design complexo [3](#)
amostragem aleatória simples
 no Assistente de amostragem [4](#)
amostragem complexa
 plano de amostra [3](#)
 plano de análise [8](#)
amostragem PPS
 no Assistente de amostragem [4](#)
amostragem sequencial
 no Assistente de amostragem [4](#)
amostragem sistemática
 no Assistente de amostragem [4](#)
Amostras complexas
 opções [13](#), [15](#), [17](#), [18](#)
 testes de hipótese [20](#), [24](#), [29](#)
 valores omissos [13](#), [16](#)
arquivo de plano [2](#)

B

Bonferroni
 em Amostras complexas [20](#), [24](#), [29](#)
 na Regressão de Cox de amostras complexas [36](#)

C

camada de linha de base
 na Regressão de Cox de amostras complexas [34](#)
categoria de referência
 na Regressão logística de amostras complexas [23](#)
 no Modelo linear geral de amostras complexas [21](#)
categorias previstas
 na Regressão logística de amostras complexas [25](#)
 na Regressão ordinal de amostras complexas [30](#)
clusters
 no Assistente de amostragem [3](#)
 no Assistente de preparação de análise [9](#)
coeficiente de variação (COV)
 em Descritivos de Amostras Complexas [14](#)
 em Frequências de amostras complexas [12](#)
 em Razões de amostras complexas [18](#)
 nas Tabulações cruzadas de amostras complexas [15](#)
contagem não ponderada
 em Descritivos de Amostras Complexas [14](#)
 em Frequências de amostras complexas [12](#)
 em Razões de amostras complexas [18](#)
 nas Tabulações cruzadas de amostras complexas [15](#)
contrastes
 no Modelo linear geral de amostras complexas [21](#)
contrastes de desvio
 no Modelo linear geral de amostras complexas [21](#)
contrastes de diferença
 no Modelo linear geral de amostras complexas [21](#)
contrastes de Helmert

 contrastes de Helmert (*continuação*)
 no Modelo linear geral de amostras complexas [21](#)
contrastes polinomiais
 no Modelo linear geral de amostras complexas [21](#)
contrastes repetidos
 no Modelo linear geral de amostras complexas [21](#)
contrastes simples
 no Modelo linear geral de amostras complexas [21](#)
convergência de parâmetro
 na Regressão logística de amostras complexas [26](#)
 na Regressão ordinal de amostras complexas [30](#)
convergência de probabilidade
 na Regressão logística de amostras complexas [26](#)
 na Regressão ordinal de amostras complexas [30](#)
correção de Bonferroni sequencial
 em Amostras complexas [20](#), [24](#), [29](#)
 na Regressão de Cox de amostras complexas [36](#)
correção de Sidak
 em Amostras complexas [20](#), [24](#), [29](#)
correção de Sidak sequencial
 em Amostras complexas [20](#), [24](#), [29](#)
 na Regressão de Cox de amostras complexas [36](#)
correção Sidak
 na Regressão de Cox de amostras complexas [36](#)
correlações de estimativas paramétrica
 na Regressão logística de amostras complexas [24](#)
 na Regressão ordinal de amostras complexas [28](#)
 no Modelo linear geral de amostras complexas [20](#)
covariâncias de estimativas paramétrica
 na Regressão logística de amostras complexas [24](#)
 na Regressão ordinal de amostras complexas [28](#)
 no Modelo linear geral de amostras complexas [20](#)

D

Descrições de amostras complexas
 estatísticas [14](#)
 valores omissos [14](#)
diferença de risco
 nas Tabulações cruzadas de amostras complexas [15](#)
diferença menos significativa
 em Amostras complexas [20](#), [24](#), [29](#)
 na Regressão de Cox de amostras complexas [36](#)

E

efeito do design
 em Descritivos de Amostras Complexas [14](#)
 em Frequências de amostras complexas [12](#)
 em Razões de amostras complexas [18](#)
 na Regressão de Cox de amostras complexas [35](#)
 na Regressão logística de amostras complexas [24](#)
 na Regressão ordinal de amostras complexas [28](#)
 nas Tabulações cruzadas de amostras complexas [15](#)
 no Modelo linear geral de amostras complexas [20](#)
erro padrão
 em Descritivos de Amostras Complexas [14](#)

erro padrão (*continuação*)
em Frequências de amostras complexas [12](#)
em Razões de amostras complexas [18](#)
na Regressão logística de amostras complexas [24](#)
na Regressão ordinal de amostras complexas [28](#)
nas Tabulações cruzadas de amostras complexas [15](#)
no Modelo linear geral de amostras complexas [20](#)

Escoragem de Fisher
na Regressão ordinal de amostras complexas [30](#)

estatística F
em Amostras complexas [20](#), [24](#), [29](#)
na Regressão de Cox de amostras complexas [36](#)

estatística F ajustada
em Amostras complexas [20](#), [24](#), [29](#)
na Regressão de Cox de amostras complexas [36](#)

estatística R2
no Modelo linear geral de amostras complexas [20](#)

estimação de amostragem
no Assistente de preparação de análise [9](#)

estimativas de parâmetro
na Regressão de Cox de amostras complexas [35](#)
na Regressão logística de amostras complexas [24](#)
na Regressão ordinal de amostras complexas [28](#)
no Modelo linear geral de amostras complexas [20](#)

estratificação
no Assistente de amostragem [3](#)
no Assistente de preparação de análise [9](#)

etapa pela metade
na Regressão logística de amostras complexas [26](#)
na Regressão ordinal de amostras complexas [30](#)

F

Frequências de amostras complexas
estatísticas [12](#)

G

graus de liberdade
em Amostras complexas [20](#), [24](#), [29](#)
na Regressão de Cox de amostras complexas [36](#)

H

histórico de iteração
na Regressão logística de amostras complexas [26](#)
na Regressão ordinal de amostras complexas [30](#)

I

informações de plano de amostra
na Regressão de Cox de amostras complexas [35](#)

intervalos de confiança
em Descritivos de Amostras Complexas [14](#)
em Frequências de amostras complexas [12](#)
em Razões de amostras complexas [18](#)
na Regressão logística de amostras complexas [24](#)
na Regressão ordinal de amostras complexas [28](#)
nas Tabulações cruzadas de amostras complexas [15](#)
no Modelo linear geral de amostras complexas [20](#), [22](#)

iterações
na Regressão logística de amostras complexas [26](#)
na Regressão ordinal de amostras complexas [30](#)

M

média
em Descritivos de Amostras Complexas [14](#)

médias marginais estimadas
no Modelo linear geral de amostras complexas [21](#)

medida de tamanho
no Assistente de amostragem [4](#)

método de amostragem
no Assistente de amostragem [4](#)

método de amostragem de Brewer
no Assistente de amostragem [4](#)

método de amostragem de Murthy
no Assistente de amostragem [4](#)

método de amostragem de Sampford
no Assistente de amostragem [4](#)

Método de estimação de Breslow
na Regressão de Cox de amostras complexas [39](#)

Método de estimação de Efron
na Regressão de Cox de amostras complexas [39](#)

método Newton-Raphson
na Regressão ordinal de amostras complexas [30](#)

Modelo linear geral de amostras complexas
estatísticas [20](#)
médias estimadas [21](#)
modelo [19](#)
opções [22](#)
recursos adicionais do comando [22](#)
salvar variáveis [21](#)

N

nível de confiança
na Regressão logística de amostras complexas [26](#)
na Regressão ordinal de amostras complexas [30](#)

P

pesos de amostra
no Assistente de amostragem [6](#)
no Assistente de preparação de análise [9](#)

plano de amostra [3](#)

plano de análise [8](#)

ponderações de amostra de entrada
no Assistente de amostragem [3](#)

porcentagens da tabela
em Frequências de amostras complexas [12](#)
nas Tabulações cruzadas de amostras complexas [15](#)

porcentagens de coluna
nas Tabulações cruzadas de amostras complexas [15](#)

porcentagens de linha
nas Tabulações cruzadas de amostras complexas [15](#)

preditor dependente de tempo
na Regressão de Cox de amostras complexas [34](#)

probabilidade predita
na Regressão logística de amostras complexas [25](#)
na Regressão ordinal de amostras complexas [30](#)

probabilidades acumulativas
na Regressão ordinal de amostras complexas [30](#)

probabilidades de inclusão
no Assistente de amostragem [6](#)

probabilidades de resposta
na Regressão ordinal de amostras complexas [27](#)

proporção de amostra
no Assistente de amostragem [6](#)
pseudoestatísticas R2
na Regressão logística de amostras complexas [24](#)
na Regressão ordinal de amostras complexas [28](#)

Q

qui-quadrado
em Amostras complexas [20](#), [24](#), [29](#)
na Regressão de Cox de amostras complexas [36](#)
qui-quadrado ajustado
em Amostras complexas [20](#), [24](#), [29](#)
na Regressão de Cox de amostras complexas [36](#)

R

raiz quadrada do efeito de design
em Descritivos de Amostras Complexas [14](#)
em Frequências de amostras complexas [12](#)
em Razões de amostras complexas [18](#)
na Regressão de Cox de amostras complexas [35](#)
na Regressão logística de amostras complexas [24](#)
na Regressão ordinal de amostras complexas [28](#)
nas Tabulações cruzadas de amostras complexas [15](#)
no Modelo linear geral de amostras complexas [20](#)
Razões de amostras complexas
estatísticas [18](#)
valores omissos [18](#)
razões de chances
na Regressão logística de amostras complexas [25](#)
na Regressão ordinal de amostras complexas [30](#)
nas Tabulações cruzadas de amostras complexas [15](#)
Regressão de Cox de amostras complexas
análise de Kaplan-Meier [31](#)
definir evento [33](#)
estatísticas [35](#)
exportação de modelo [38](#)
modelo [34](#)
opções [39](#)
plots [36](#)
preditor dependente de tempo [34](#)
preditores [33](#)
salvar variáveis [37](#)
subgrupos [34](#)
testes de hipótese [36](#)
variáveis de data e hora [31](#)
Regressão logística de amostras complexas
categoria de referência [23](#)
estatísticas [24](#)
modelo [23](#)
opções [26](#)
razões de chances [25](#)
recursos adicionais do comando [26](#)
salvar variáveis [25](#)
Regressão ordinal de amostras complexas
estatísticas [28](#)
modelo [28](#)
opções [30](#)
probabilidades de resposta [27](#)
razões de chances [30](#)
salvar variáveis [30](#)
resíduos

resíduos (*continuação*)
nas Tabulações cruzadas de amostras complexas [15](#)
no Modelo linear geral de amostras complexas [21](#)
resíduos agregados
na Regressão de Cox de amostras complexas [37](#)
resíduos ajustados
nas Tabulações cruzadas de amostras complexas [15](#)
resíduos de Cox-Snell
na Regressão de Cox de amostras complexas [37](#)
resíduos de deviance
na Regressão de Cox de amostras complexas [37](#)
resíduos de escore
na Regressão de Cox de amostras complexas [37](#)
resíduos de martingale
na Regressão de Cox de amostras complexas [37](#)
Resíduos parciais de Schoenfeld
na Regressão de Cox de amostras complexas [37](#)
risco relativo
nas Tabulações cruzadas de amostras complexas [15](#)

S

separação
na Regressão logística de amostras complexas [26](#)
na Regressão ordinal de amostras complexas [30](#)
soma
em Descritivos de Amostras Complexas [14](#)
subpopulação
na Regressão de Cox de amostras complexas [34](#)

T

tabelas de classificação
na Regressão logística de amostras complexas [24](#)
na Regressão ordinal de amostras complexas [28](#)
Tabulações cruzadas de amostras complexas
estatísticas [15](#)
tamanho da amostra
no Assistente de amostragem [5](#), [6](#)
tamanho da população
em Descritivos de Amostras Complexas [14](#)
em Frequências de amostras complexas [12](#)
em Razões de amostras complexas [18](#)
nas Tabulações cruzadas de amostras complexas [15](#)
no Assistente de amostragem [6](#)
teste de linhas paralelas
na Regressão ordinal de amostras complexas [28](#)
teste de riscos proporcionais
na Regressão de Cox de amostras complexas [35](#)
Teste t
na Regressão logística de amostras complexas [24](#)
na Regressão ordinal de amostras complexas [28](#)
no Modelo linear geral de amostras complexas [20](#)

V

valores acumulativos
em Frequências de amostras complexas [12](#)
valores esperados
nas Tabulações cruzadas de amostras complexas [15](#)
valores omissos
em Amostras complexas [13](#), [16](#)
em Descritivos de Amostras Complexas [14](#)

valores omissos (*continuação*)
em Razões de amostras complexas 18
na Regressão logística de amostras complexas 26
na Regressão ordinal de amostras complexas 30
no Modelo linear geral de amostras complexas 22
valores preditos
no Modelo linear geral de amostras complexas 21

