

IBM SPSS Custom Tables 25

IBM

Comunicado

Antes de usar estas informações e o produto suportado por elas, leia as informações nos “Avisos” na página 89.

Informações sobre o produto

Esta edição aplica-se à versão 25, liberação 0, modificação 0 do IBM SPSS Statistics e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

Índice

Tabelas customizadas.	1
Interface do construtor de tabela.	1
Interface do construtor de tabela.	1
Construindo tabelas	1
Tabelas customizadas: Guia Opções	16
Tabelas customizadas: Guia Títulos	17
Tabelas customizadas: Guia Estatísticas de teste	18
Tabelas simples para variáveis categóricas	19
Tabelas simples para variáveis categóricas	19
Uma variável categórica única	20
Tabulação cruzada	22
Ordenando e excluindo categorias	24
Empilhamento, aninhamento e estratos com variáveis categóricas	25
Empilhando variáveis categóricas	26
Aninhando variáveis categóricas	27
Camadas	30
Totais e subtotais para variáveis categóricas	32
Total simples para uma variável única	32
O que você vê é o que será totalizado	32
Posição de exibição de totais.	33
Totais para tabelas aninhadas	33
Totais de variáveis de estrato	35
Subtotais	35
Categorias calculadas para variáveis categóricas	37
Categoria calculada simples	37
Ocultando categorias em uma categoria calculada	38
Referenciando subtotais em uma categoria calculada	39
Usando categorias calculadas para exibir subtotais não exaustivos	40
Tabelas para variáveis com categorias compartilhadas	41
Tabela de contagens	42
Tabela de porcentagens	42
Controle de totais e de categorias	43
Aninhando em tabelas com categorias compartilhadas	44
Estatísticas básicas	44
Variável de origem estatística de sumarização	45
Variáveis empilhadas	47
Total de estatísticas básicas customizada para variáveis categóricas	47
Sumarizando variáveis de escala	49
Sumarizando variáveis de escala	49
Variáveis de escala empilhadas	50
Várias estatísticas básicas	50
Contagem, N válido e Valores omissos	51
Sumarizações diferentes para variáveis diferentes	52
Agrupar sumarizações em categorias	52
Intervalos de confiança	54
Estatísticas de teste	55
Estatísticas de teste	55
Testes de independência (Qui-quadrado)	55
Comparando médias de coluna.	58
Comparando proporções da coluna	61
Uma nota sobre ponderações e conjuntos de múltiplas respostas	66
Conjuntos de múltiplas respostas	66
Contagens, respostas, porcentagens e totais.	67
Usando conjuntos de múltiplas respostas com outras variáveis	69
Conjuntos de categorias múltiplas e respostas duplicadas.	70
Teste de significância com conjuntos de múltiplas respostas	72
Valores omissos	73
Tabelas sem valores omissos.	74
Incluindo valores omissos em tabelas.	74
Formatando e customizando tabelas	76
Formatando e customizando tabelas	76
Formato de exibição de estatísticas básicas	76
Rótulos de exibição para estatísticas básicas	77
Largura da coluna	78
Valor de exibição para células vazias	78
Arquivos de Amostra	80
Avisos	89
Marcas comerciais	91
Índice Remissivo	93

Tabelas customizadas

Os recursos de tabelas customizadas a seguir estão incluídos em SPSS Statistics Standard Edition ou a opção Tabelas customizadas.

Interface do construtor de tabela

Interface do construtor de tabela

As Tabelas customizadas usam uma interface simples do construtor de tabela arrastar e soltar que permite visualizar sua tabela conforme você seleciona variáveis e opções. Também fornecem um nível de flexibilidade não localizado em uma caixa de diálogo típica, incluindo a capacidade de mudar o tamanho da janela e o tamanho das áreas de janela na janela.

Construindo tabelas

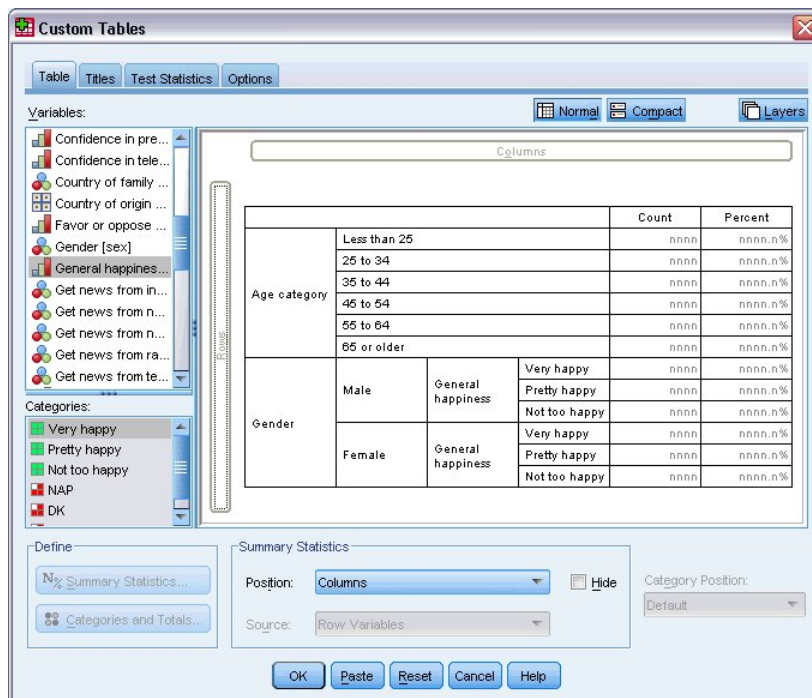


Figura 1. Caixa de diálogo Tabelas customizadas, guia Tabela

Selecione as variáveis e medidas de sumarização que aparecerão em suas tabelas na guia Tabela no construtor de tabela.

Lista de variáveis. As variáveis no arquivo de dados são exibidas na área de janela superior esquerda da janela. As Tabelas customizadas distinguem entre dois diferentes níveis de medição para variáveis e trata-as de forma diferente, dependendo do nível de medição:

Catagórico. Dados com um número limitado de valores ou categorias distintas (por exemplo, sexo ou religião). Variáveis categóricas podem ser variáveis de sequência de caracteres (alfanuméricas) ou

variáveis numéricas que usam códigos numéricos para representar categorias (por exemplo, 0 = *male* e 1 = *female*). Também referidos como dados qualitativos. Variáveis categóricas podem ser **nominais** ou **ordinais**

- *Nominal*. Uma variável pode ser tratada como nominal quando seus valores representarem categorias sem ranqueamento intrínseco (por exemplo, o departamento da empresa na qual um funcionário trabalha). Exemplos de variáveis nominais incluem região, código de endereçamento postal e filiação religiosa.
- *Ordinal*. Uma variável pode ser tratada como ordinal quando seus valores representarem categorias com algum ranqueamento intrínseco (por exemplo, níveis de satisfação de serviço de muito insatisfeito para muito satisfeito). Exemplos de variáveis ordinais incluem escores de atitude que representam o grau de satisfação ou de confiança e os escores de classificação de preferência.

Escala. Dados medidos em um intervalo ou razão de escala, em que os valores de dados indicam a ordem dos valores e a distância entre os valores. Por exemplo, um salário de \$72.195 é maior que um salário de \$52.398, e a distância entre os dois valores é de \$19.797. Também é conhecida como dados quantitativos ou contínuos.

As variáveis categóricas definem categorias (linha, colunas e camadas) na tabela e a estatística de sumarização padrão é a contagem (número de casos em cada categoria). Por exemplo, uma tabela padrão de variável de gênero categórico simplesmente exibiria o número de homens e o número de mulheres.

As variáveis de escala geralmente são sumarizadas em categorias de variáveis categóricas e a estatística de sumarização padrão é a média. Por exemplo, uma tabela padrão de receita em categorias de gênero exibiria a receita média para homens e a receita média para mulheres.

Também é possível sumarizar variáveis de escala por elas mesmas, sem usar uma variável categórica para definir grupos. Isso é útil principalmente para **empilhar** sumarizações de diversas variáveis de escala. Consulte o tópico “Empilhando variáveis” na página 4 para obter mais informações

Conjuntos de múltiplas respostas

As Tabelas customizadas também suportam um tipo especial de "variável" chamado **conjunto de múltiplas respostas**. Os conjuntos de múltiplas respostas não são realmente variáveis no sentido normal. Não é possível vê-los no Editor de dados e outros procedimentos não os reconhecem. Os conjuntos de múltiplas respostas usam variáveis múltiplas para registrar respostas a perguntas em que o respondente pode dar mais de uma resposta. Os conjuntos de múltiplas respostas são tratados como variáveis categóricas e a maioria das coisas que podem ser feitas com variáveis categóricas também podem ser feitas com conjuntos de múltiplas respostas. Consulte o tópico “Conjuntos de múltiplas respostas” na página 66 para obter mais informações

Um ícone próximo a cada variável na lista de variáveis identifica o tipo de variável.

Tabela 1. Ícones do nível de medição








	Numérico	Sequência de caracteres	Data	Horário
Escala (Contínua)		n/a		
Ordinal				

Tabela 1. Ícones do nível de medição (continuação)







	Numérico	Sequência de caracteres	Data	Horário
Nominal				

Tabela 2. Ícones de conjunto de múltiplas respostas

Tipo de conjunto de múltiplas respostas	Ícone
Conjunto de múltiplas respostas, diversas categorias	
Conjunto de múltiplas respostas, diversas dicotomias	

É possível mudar o nível de medição de uma variável no construtor de tabela, clicando com o botão direito na variável na lista de variáveis e selecionando **Catagórico** ou **Escala** no menu pop-up. É possível mudar permanentemente o nível de medição de uma variável na visualização Variável do Editor de dados. As variáveis definidas como **nominais** ou **ordinais** são tratadas como categóricas por Tabelas customizadas.

Categorias. Ao selecionar uma variável categórica na lista de variáveis, as categorias definidas para a variável são exibidas na lista Categorias. Essas categorias também serão exibidas na área de janela de tela quando a variável for usada em uma tabela. Se a variável não tiver nenhuma categoria definida, a lista Categorias e a área de janela de tela exibirão duas categorias de item temporário: *Categoria 1* e *Categoria 2*.

As categorias definidas exibidas no construtor de tabela são baseadas em **rótulos de valor**, rótulos descritivos designados a diferentes valores de dados (por exemplo, valores numéricos de 0 e 1, com rótulos de valor de *homem* e *mulher*). É possível definir rótulos de valor na visualização Variável do Editor de dados ou com Definir propriedades da variável no menu Dados na janela Editor de dados.

Área de janela de tela. Construa uma tabela arrastando e soltando variáveis nas linhas e colunas da área de janela de tela. A área de janela de tela exibe uma visualização da tabela que será criada. A área de janela de tela não mostra valores de dados reais nas células, mas deve fornecer uma visualização bastante precisa do layout da tabela final. Para variáveis categóricas, a tabela real pode conter mais categorias que a visualização se o arquivo de dados contiver valores exclusivos para os quais nenhum rótulo de valor foi definido.

- A visualização **Normal** exibe todas as linhas e colunas que serão incluídas na tabela, incluindo linhas e/ou colunas para estatísticas básicas e categorias de variáveis categóricas.
- A visualização **Compacta** mostra apenas as variáveis que estarão na tabela, sem uma visualização das linhas e colunas que a tabela conterá.

Regras básicas e limitações para construir uma tabela

- Para variáveis categóricas, as estatísticas básicas são baseadas na variável mais interna na dimensão de origem estatística.
- A dimensão de origem estatística padrão (linha ou coluna) para variáveis categóricas é baseada na ordem em que você arrasta e solta variáveis na área de janela de tela. Por exemplo, se você arrastar uma variável para a bandeja de linhas primeiro, a dimensão da linha será a dimensão de origem estatística padrão.

- As variáveis de escala podem ser sumarizadas apenas nas categorias da variável mais interna na dimensão da linha ou da coluna. (É possível posicionar a variável de escala em qualquer nível da tabela, mas ela é sumarizada no nível mais interno.)
- As variáveis de escala não podem ser sumarizadas em outras variáveis de escala. É possível empilhar sumarizações de variáveis de escala múltiplas ou sumarizar variáveis de escala em categorias de variáveis categóricas. Não é possível aninhar uma variável de escala em outra ou colocar uma variável de escala na dimensão da linha e outra variável de escala na dimensão da coluna.
- Se alguma variável no conjunto de dados ativo contiver mais de 12.000 rótulos de valor definidos, não será possível usar o construtor de tabela para criar tabelas. Se não precisar incluir variáveis que excedem essa limitação em suas tabelas, é possível definir e aplicar conjuntos de variáveis que excluem essas variáveis. Se precisar incluir variáveis com mais de 12.000 rótulos de valor definidos, é possível usar a sintaxe de comando CTABLES para gerar as tabelas.

Para construir uma tabela

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. Arraste e solte uma ou mais variáveis para as áreas de linha e/ou coluna da área de janela de tela.
3. Clique em **OK** para criar a tabela.
4. Selecione (clique) a variável na área de janela de tela.
5. Arraste a variável para qualquer lugar fora da área de janela de tela ou pressione a tecla Delete.
Para mudar o nível de medição de uma variável:
6. Clique com o botão direito na variável na lista de variáveis (isso pode ser feito apenas na lista de variáveis, não na tela).
7. Selecione **Catagórico** ou **Escala** no menu pop-up.

Empilhando variáveis

O empilhamento pode ser considerado como pegar tabelas separadas e colá-las juntas na mesma exibição. Por exemplo, é possível exibir informações sobre *Gênero* e *Categoria de idade* em seções separadas da mesma tabela.

Para empilhar variáveis

1. Na lista de variáveis, selecione todas as variáveis que você deseja empilhar, em seguida, arraste e solte-as juntas nas linhas ou colunas da área de janela de tela.
ou
2. Arraste e solte variáveis separadamente, soltando cada variável acima ou abaixo de variáveis existentes nas linhas ou à direita ou esquerda de variáveis existentes nas colunas.

Tabela 3. Variáveis categóricas empilhadas.

Variáveis	Categorias	Estatística de resumo
Variável 1	Categoria 1	123
	Categoria 2	456
Variável 2	Categoria 1	123
	Categoria 2	456
	Categoria 3	789

Consulte o tópico “Empilhando variáveis categóricas” na página 26 para obter mais informações

Aninhando variáveis

O aninhamento, assim como a tabulação cruzada, pode mostrar o relacionamento entre duas variáveis categóricas, exceto que uma variável está aninhada na outra na mesma dimensão. Por exemplo, é possível aninhar *Gênero* em *Categoria de idade* na dimensão da linha, mostrando o número de homens e mulheres em cada categoria de idade.

Também é possível aninhar uma variável de escala em uma variável categórica. Por exemplo, é possível aninhar *Receita* em *Gênero*, mostrando valores médios (ou mediana ou outra medida de sumarização) de receita separados para homens e mulheres.

Para aninhar variáveis

1. Arraste e solte uma variável categórica na área de linha ou de coluna da área de janela de tela.
2. Arraste e solte uma variável categórica ou de escala à esquerda ou à direita da variável de linha categórica ou acima ou abaixo da variável de coluna categórica.

Tabela 4. Variáveis categóricas aninhadas.

Variável 1	Variável 2	Estatística de resumo
Categoria 1	Categoria 1	12
	Categoria 2	34
	Categoria 3	56
Categoria 2	Categoria 1	12
	Categoria 2	34
	Categoria 3	56

Consulte o tópico “Aninhando variáveis categóricas” na página 27 para obter mais informações

Nota: As Tabelas customizadas não consideram o processamento de arquivo dividido em estratos. Para obter o mesmo resultado como arquivos divididos em estratos, coloque as variáveis de arquivo dividido nas estratos de aninhamento mais externas da tabela.

Camadas

É possível usar estratos para incluir uma dimensão de espessura em suas tabelas, criando "cubos tridimensionais." As estratos são semelhantes ao aninhamento ou empilhamento; a principal diferença é que somente uma categoria de estrato fica visível de cada vez. Por exemplo, usar *Categoria de idade* como a variável de linha e *Gênero* como uma variável de estrato produz uma tabela na qual as informações para homens e mulheres são exibidas em diferentes estratos da tabela.

Para criar estratos

1. Clique em **Camadas** na guia Tabela no construtor de tabela para exibir a lista Camadas.
2. Arraste e solte as variáveis de escala ou categóricas que definirão as estratos para a lista Camadas.

Não é possível combinar variáveis de escala e categóricas na lista Camadas. Todas as variáveis devem ser do mesmo tipo. Os conjuntos de múltiplas respostas são tratados como categóricos para a lista Camadas. As variáveis de escala nas estratos são sempre empilhadas.

Se você tiver diversas variáveis de estrato categóricas, as estratos podem ser empilhadas ou aninhadas.

- **Mostrar cada categoria como uma estrato** é equivalente ao empilhamento. Uma estrato separada será exibida para cada categoria de cada variável de estrato. O número total de estratos é simplesmente a soma do número de categorias para cada variável de estrato. Por exemplo, se você tiver três variáveis de estrato, cada uma com três categorias, a tabela terá nove estratos.

- **Mostrar cada combinação de categorias como uma estrato** é equivalente ao aninhamento ou tabulação cruzada de estratos. O número total de estratos é o *produto* do número de categorias para cada variável de estrato. Por exemplo, se você tiver três variáveis, cada uma com três categorias, a tabela terá 27 estratos.

Mostrando e ocultando nomes e/ou rótulos de variáveis

As seguintes opções estão disponíveis para a exibição de nomes e rótulos de variáveis:

- Mostrar apenas rótulos de variáveis. Para variáveis sem rótulos de variáveis definidos, o nome de variável é exibido. Esta é a definição padrão.
- Mostrar apenas nomes de variáveis.
- Mostrar rótulos de variáveis e nomes de variáveis.
- Não mostrar nomes de variáveis ou rótulos de variáveis. Embora a coluna/linha que contém o rótulo ou nome da variável ainda será exibida na visualização da tabela na área de janela de tela, essa coluna/linha não será exibida na tabela real.

Para mostrar ou ocultar rótulos de variáveis ou nomes de variáveis:

1. Clique com o botão direito na variável na visualização da tabela na área de janela de tela.
2. Selecione **Mostrar rótulo de variáveis** ou **Mostrar nome de variável** no menu pop-up para ativar ou desativar a exibição de rótulos ou nomes. Uma marca de seleção próxima da seleção indica que ela será exibida.

Estatísticas básicas

A caixa de diálogo estatísticas básicas permite:

- Incluir e remover estatísticas básicas de uma tabela.
- Mudar os rótulos para as estatísticas.
- Mudar a ordem das estatísticas.
- Mudar o formato das estatísticas, incluindo o número de casas decimais.

As estatísticas básicas (e outras opções) disponíveis aqui dependem do nível de medição da variável de origem estatística de sumarização, conforme exibido na parte superior da caixa de diálogo. A origem das estatísticas básicas (a variável na qual as estatísticas básicas são baseadas) é determinada por:

- **Nível de medição.** Se uma tabela (ou uma seção da tabela em uma tabela empilhada) contiver uma variável de escala, as estatísticas básicas serão baseadas na variável de escala.
- **Ordem de seleção de variáveis.** A dimensão de origem estatística padrão (linha ou coluna) para variáveis categóricas é baseada na ordem em que você arrasta e solta variáveis na área de janela de tela. Por exemplo, se você arrastar uma variável para a área de linhas primeiro, a dimensão da linha será a dimensão de origem estatística padrão.
- **Aninhamento.** Para variáveis categóricas, as estatísticas básicas são baseadas na variável mais interna na dimensão de origem estatística.

Uma tabela empilhada pode ter diversas variáveis de origem estatística de sumarização (de escala e categóricas), mas cada seção da tabela possui apenas uma origem estatística de sumarização.

Para mudar a dimensão de origem estatística de sumarização

1. Selecione a dimensão (linhas, colunas ou camadas) da lista suspensa **Origem** no grupo estatísticas básicas da guia Tabela.

Para controlar as estatísticas básicas exibidas em uma tabela

1. Selecione (clique) a variável de origem estatística de sumarização na área de janela de tela da guia Tabela.
2. No grupo Definição da guia Tabela, clique em **estatísticas básicas**.

ou

3. Clique com o botão direito na variável de origem estatística de sumarização na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
4. Selecione as estatísticas básicas que você deseja incluir na tabela. É possível usar a seta para mover estatísticas selecionadas da lista Estatísticas para a lista Exibição, ou é possível arrastar e soltar estatísticas selecionadas da lista Estatísticas para a lista Exibição.
5. Clique nas setas para cima ou para baixo para mudar a posição de exibição da estatística de sumarização atualmente selecionada.
6. Selecione um formato de exibição da lista suspensa Formato para a estatística de sumarização selecionada.
7. Insira o número de casas decimais a serem exibidas na célula Casas decimais para a estatística de sumarização selecionada.
8. Clique em **Aplicar à seleção** para incluir as estatísticas básicas selecionadas para as variáveis selecionadas atualmente na área de janela de tela.
9. Clique em **Aplicar a todos** para incluir as estatísticas básicas selecionadas para todas as variáveis empilhadas do mesmo tipo na área de janela de tela.

Nota: **Aplicar a todos** difere de **Aplicar à seleção** apenas para variáveis empilhadas do mesmo tipo já na área de janela de tela. Em ambos os casos, as estatísticas básicas selecionadas são incluídas automaticamente para qualquer variável empilhada adicional do mesmo tipo que for incluída na tabela.

estatísticas básicas para variáveis categóricas: As estatísticas básicas disponíveis para variáveis categóricas são contagens e porcentagens. Também é possível especificar estatísticas básicas customizada para totais e subtotais. Essas estatísticas básicas customizada incluem medidas de tendência central (como média e mediana) e dispersão (como desvio padrão) que podem ser adequadas para algumas variáveis categóricas ordinais. Consulte o tópico “Total de estatísticas básicas customizada para variáveis categóricas” na página 10 para obter mais informações

Contagem. Número de casos em cada célula da tabela ou número de respostas para conjuntos de múltiplas respostas. Se peso estiver em vigor, esse valor é a contagem ponderada.

- Se peso estiver em vigor, o valor é a contagem ponderada.
- A contagem ponderada é a mesma para ambos: o peso do conjunto de dados global (**Dados > Casos de Ponderação**) e o peso com uma variável de ponderação de base efetiva especificada na guia Opções do diálogo Tabelas Customizadas.

Contagem não ponderada. Número não ponderado de casos em cada célula da tabela. Isso difere da contagem apenas se a ponderação estiver em vigor.

Contagem Ajustada. A contagem ajustada usada em cálculos de ponderação de base efetiva. Se você não usar uma variável de ponderação de base efetiva (guia Opções), a contagem ajustada é a mesma que a contagem.

Porcentagens da coluna. Porcentagens em cada coluna. As porcentagens em cada coluna de uma subtabela (para porcentagens simples) somam 100%. As porcentagens da coluna geralmente serão úteis apenas se você tiver uma variável de *linha* categórica.

Porcentagens de linhas. Porcentagens em cada linha. As porcentagens em cada linha de uma subtabela (para porcentagens simples) somam 100%. As porcentagens de linhas geralmente serão úteis apenas se você tiver uma variável de *coluna* categórica.

Porcentagens de linhas da estrato e de colunas da estrato. As porcentagens de linhas ou de colunas (para porcentagens simples) somam 100% em todas as subtabelas em uma tabela aninhada. Se a tabela contiver estratos, as porcentagens de linhas ou de colunas somarão 100% em todas as subtabelas aninhadas em cada estrato.

Porcentagens de estratos. Porcentagens em cada estrato. Para porcentagens simples, as porcentagens de células na estrato visível atualmente somam 100%. Se você não tiver nenhuma variável de estrato, isso será equivalente às porcentagens da tabela.

Porcentagens da tabela. As porcentagens para cada célula são baseadas na tabela inteira. Todas as porcentagens de células são baseadas no mesmo número total de casos e somam 100% (para porcentagens simples) na tabela inteira.

Porcentagens da subtabela. As porcentagens em cada célula são baseadas na subtabela. Todas as porcentagens de células na subtabela são baseadas no mesmo número total de casos e somam 100% na subtabela. Em tabelas aninhadas, a variável que precede o nível de aninhamento mais interno define as subtabelas. Por exemplo, em uma tabela de *Estado civil em Gênero* na *Categoria de idade, Gênero* define subtabelas.

Intervalos de confiança

- Os limites de confiança inferior e superior estão disponíveis para contagens, porcentagens, média, mediana, percentis e soma.
- A sequência de caracteres de texto "&[Nível de confiança]" no rótulo inclui o nível de confiança no rótulo da coluna na tabela.
- O erro padrão está disponível para contagens, porcentagens, média e soma.
- Os intervalos de confiança e o erro padrão não estão disponíveis para conjuntos de múltiplas respostas.

Nível O nível de confiança para intervalos de confiança, expresso como uma porcentagem. O valor deve ser maior que 0 e menor que 100.

Conjuntos de múltiplas respostas

Os conjuntos de múltiplas respostas podem ter porcentagens baseadas em casos, respostas ou contagens. Consulte o tópico "estatísticas básicas para conjuntos de múltiplas respostas" na página 9 para obter mais informações

Tabelas empilhadas

Para cálculos de porcentagem, cada seção da tabela definida por uma variável de empilhamento é tratada como uma tabela separada. As porcentagens de Linhas da estrato, de Colunas da estrato e da Tabela somam 100% (para porcentagens simples) em cada seção da tabela empilhada. A base de porcentagem para diferentes cálculos de porcentagem é baseada nos casos em cada seção da tabela empilhada.

Base de porcentagem: As porcentagens podem ser calculadas de três maneiras diferentes, determinadas pelo tratamento de valores omissos na base computacional:

Porcentagem simples. As porcentagens são baseadas no número de casos usados na tabela e sempre somam 100%. Se uma categoria for excluída da tabela, os casos nessa categoria são excluídos da base. Casos com valores omissos do sistema são sempre excluídos da base. Os casos com valores omissos de usuário serão excluídos se as categorias com usuário desconhecido forem excluídas da tabela (o padrão) ou incluídos se as categorias com usuário desconhecido forem incluídas na tabela. Qualquer porcentagem que não tenha *N válido* ou *N total* em seu nome é uma porcentagem simples.

Porcentagem de N total. Os casos com valores omissos do sistema e omissos de usuário são incluídos na base de porcentagem Simples. As porcentagens podem somar menos de 100%.

Porcentagem de N válido. Os casos com valores omissos de usuário são removidos da base de porcentagem Simples, mesmo que as categorias com usuário desconhecido sejam incluídas na tabela.

Nota: Os casos em categorias excluídas manualmente diferentes de categorias com usuário desconhecido são sempre excluídos da base.

estatísticas básicas para conjuntos de múltiplas respostas: As seguintes estatísticas básicas adicionais estão disponíveis para conjuntos de múltiplas respostas.

% de Respostas de Col/Linha/Camada. Porcentagem baseada em respostas.

% de Respostas de Col/Linha/Camada (Base: Contagem). As respostas são o numerador e a contagem total é o denominador.

% de Contagem de Col/Linha/Camada (Base: Respostas). Contagem é o numerador e total de respostas é o denominador.

% de Respostas de Col/Linha de estrato. Porcentagem em subtabelas. Porcentagem baseada em respostas.

% de Respostas de Col/Linha de estrato (Base: Contagem). Porcentagens em subtabelas. As respostas são o numerador e a contagem total é o denominador.

% de Respostas de Col/Linha de estrato (Base: Respostas). Porcentagens em subtabelas. Contagem é o numerador e total de respostas é o denominador.

Respostas. Contagem de respostas.

% de Respostas de Subtabela/Tabela. Porcentagem baseada em respostas.

% de Respostas de Subtabela/Tabela (Base: Contagem). As respostas são o numerador e a contagem total é o denominador.

% de Contagem de Subtabela/Tabela (Base: Respostas). Contagem é o numerador e total de respostas é o denominador.

estatísticas básicas para variáveis de escala e totais customizados categóricos: Além das contagens e porcentagens disponíveis para variáveis categóricas, as seguintes estatísticas básicas estão disponíveis para variáveis de escala e como sumarizações de total e subtotal customizadas para variáveis categóricas. Essas estatísticas básicas não estão disponíveis para conjuntos de múltiplas respostas ou variáveis de sequência de caracteres (alfanuméricas).

Média. Média aritmética; a soma dividida pelo número de casos.

Mediana. Valor acima e abaixo do qual está a metade dos casos; o 50º percentil.

Modo. Valor mais frequente. Se houver um empate, o menor valor será mostrado.

Mínimo. Menor (mais baixo) valor.

Máxima. Maior (mais alto) valor.

Omisso. Contagem de valores omissos (omissos de usuário e do sistema).

Percentil. É possível incluir o 5º, 25º, 75º, 95º e/ou 99º percentis.

Intervalo. Diferença entre os valores máximo e mínimo.

Desvio padrão. Medida de dispersão ao redor da média. Em uma distribuição normal, 68% dos casos estão contidos em um desvio padrão da média e 95% dos casos estão contidos em dois desvios padrão. Por exemplo, se a média de idade for 45, com um desvio padrão de 10, 95% dos casos estariam entre 25 e 65 em uma distribuição normal (a raiz quadrada da variância).

Soma. Soma dos valores.

Porcentagem da soma. Porcentagens baseadas em somas. Disponível para linhas e colunas (em subtabelas), linhas e colunas inteiras (entre subtabelas), estratos, subtabelas e tabelas inteiras.

N total. Contagem de valores não omissos, omissos de usuário e omissos do sistema. Não inclui os casos em categorias excluídas manualmente diferentes de categorias com usuário desconhecido.

N Total Ajustado. O N total ajustado usado em cálculos de ponderação de base efetiva. Se você não usar uma variável de ponderação de base efetiva (guia Opções), o N total ajustado é o mesmo que o N total. Essa estatística não está disponível para conjuntos de múltiplas respostas.

N válido. Contagem de valores não omissos. Não inclui os casos em categorias excluídas manualmente diferentes de categorias com usuário desconhecido.

N Válido Ajustado. O em válido ajustado usado em cálculos de ponderação de base efetiva. Se você não usar uma variável de ponderação de base efetiva (guia Opções), o N válido ajustado é o mesmo que o N válido. Essa estatística não está disponível para conjuntos de múltiplas respostas.

Variância. Uma medida de dispersão em torno da média, igual à soma de desvios quadrados da média dividida por um menos o número de casos. A variância é medida em unidades que são o quadrado das unidades da própria variável (o quadrado do desvio padrão).

Intervalos de confiança

- Os limites de confiança inferior e superior estão disponíveis para contagens, porcentagens, média, mediana, percentis e soma.
- A sequência de caracteres de texto "&[Nível de confiança]" no rótulo inclui o nível de confiança no rótulo da coluna na tabela.
- O erro padrão está disponível para contagens, porcentagens, média e soma.
- Os intervalos de confiança e o erro padrão não estão disponíveis para conjuntos de múltiplas respostas.

Nível O nível de confiança para intervalos de confiança, expresso como uma porcentagem. O valor deve ser maior que 0 e menor que 100.

Tabelas empilhadas

Cada seção da tabela definida por uma variável de empilhamento é tratada como uma tabela separada, e as estatísticas básicas são calculadas de forma apropriada.

Total de estatísticas básicas customizada para variáveis categóricas: Para tabelas de variáveis categóricas que contêm totais ou subtotaís, é possível ter estatísticas básicas diferentes das sumarizações exibidas para cada categoria. Por exemplo, é possível exibir contagens e porcentagens da coluna para uma variável de linha categórica ordinal e exibir a média para a estatística "total".

Para criar uma tabela para uma variável categórica com uma estatística de sumarização total customizada:

1. Nos menus, escolha:

Analisar > Tabelas > Tabelas Customizadas...

O construtor de tabela será aberto.

2. Arraste e solte uma variável categórica na área Linhas ou Colunas da tela.
3. Clique com o botão direito na variável na tela e selecione **Categorias e totais** no menu pop-up.
4. Clique (marque) na caixa de seleção **Total** e, em seguida, clique em **Aplicar**.
5. Clique com o botão direito na variável novamente na tela e selecione **estatísticas básicas** no menu pop-up.

6. Clique em (marque) **estatísticas básicas customizada para totais e subtotais** e, em seguida, selecione as estatísticas básicas customizada desejadas.

Por padrão, todas as estatísticas básicas, incluindo as sumarizações customizadas, são exibidas na dimensão oposta da dimensão que contém a variável categórica. Por exemplo, se você tiver uma variável de linha categórica, as estatísticas básicas definem colunas na tabela, como em:

Tabela 5. Categorias de variável ordinal em linhas, contagem e média de estatísticas básicas em colunas.

Variáveis	Categorias	Contagem	Média
Variável 1	1 Concordar	196	
	2 Neutro	936	
	3 Discordar	744	
	Total	1876	

Para exibir estatísticas básicas na mesma dimensão que a variável categórica:

7. Na guia Tabela no construtor de tabela, no grupo estatísticas básicas, selecione a dimensão da lista suspensa Posição.

Por exemplo, se a variável categórica for exibida nas linhas, selecione **Linhas** na lista suspensa.

Formatos de exibição de estatísticas básicas: As seguintes opções de formato de exibição estão disponíveis:

nnnn. Numérico simples.

nnnn%. Sinal de porcentagem anexado ao final do valor.

Automático. Formato de exibição de variável definido, incluindo o número de decimais.

N=nnnn. Exibe *N=* antes do valor. Isso pode ser útil para contagens, *N* válido e *N* total nas tabelas em que os rótulos de estatísticas básicas não são exibidos.

(nnnn). Todos os valores são colocados entre parênteses.

(nnnn)(valor neg.). Apenas valores negativos colocados entre parênteses.

(nnnn%). Todos os valores colocados entre parênteses e um sinal de porcentagem anexado ao final de valores.

n,nnn.n. Formato de Vírgula. Vírgula usada como separador de agrupamento e ponto usado como indicador decimal, independentemente das configurações do código de idioma.

n.nnn,n. Formato de ponto. Ponto usado como separador de agrupamento e vírgula usada como indicador decimal, independentemente das configurações do código de idioma.

\$n,nnn.n. Formato de dólar. Símbolo de dólar exibido na frente de um valor; vírgula usada como separador de agrupamento e ponto usado como indicador decimal, independentemente das configurações do código de idioma.

CCA, CCB, CCC, CCD, CCE. Formatos de moeda customizada. O formato definido atual para cada moeda customizada é exibido na lista. Esses formatos são definidos na guia Moeda na caixa de diálogo Opções (menu Editar, Opções).

Regras gerais e limitações

- Com exceção de Automático, o número de decimais é determinado pela configuração da coluna Decimais.
- Com exceção dos formatos de vírgula, dólar e ponto, o indicador decimal usado é o definido para o código de idioma atual no painel de controle Opções regionais do Windows.
- Embora vírgula/dólar exibirão uma vírgula ou ponto respectivamente como o separador de agrupamento, não há nenhum formato de exibição disponível no horário de criação para exibir um separador de agrupamento baseado nas configurações atuais de código de idioma (definidas no painel de controle Opções regionais do Windows).

Categorias e totais

A caixa de diálogo Categorias e totais permite:

- Reordenar e excluir categorias.
- Inserir subtotais e totais.
- Inserir categorias calculadas.
- Incluir ou excluir categorias vazias.
- Incluir ou excluir categorias definidas como valores omissos de contenção.
- Incluir ou excluir categorias que não possuem rótulos de valor definido.
- Esta caixa de diálogo está disponível apenas para variáveis categóricas e conjuntos de múltiplas respostas. Ela não está disponível para variáveis de escala.
- Para variáveis múltiplas selecionadas com categorias diferentes, não é possível inserir subtotais, inserir categorias calculadas, excluir categorias ou reordenar categorias manualmente. Isso ocorre apenas se você selecionar variáveis múltiplas na visualização de tela e acessar essa caixa de diálogo para todas as variáveis selecionadas simultaneamente. Ainda é possível executar essas ações para cada variável separadamente.
- Para variáveis sem rótulos de valor definido, é possível apenas ordenar categorias e inserir totais.

Para acessar a caixa de diálogo Categorias e totais

1. Arraste e solte uma variável categórica ou conjunto de múltiplas respostas para a área de janela de tela.
2. Clique com o botão direito na variável na área de janela de tela e selecione **Categorias e totais** no menu pop-up.
ou
3. Selecione (clique) a variável na área de janela de tela e, em seguida, clique em **Categorias e totais** no grupo Definição na guia Tabela.
Também é possível selecionar variáveis categóricas múltiplas na mesma dimensão na área de janela de tela:
4. Ctrl-clique em cada variável na área de janela de tela.
ou
5. Clique fora da visualização da tabela na área de janela de tela e, em seguida, clique e arraste para selecionar a área que inclui as variáveis que você deseja selecionar.
ou
6. Clique com o botão direito em qualquer variável em uma dimensão e selecione **Selecionar todas [dimensão] as variáveis** para selecionar todas as variáveis nessa dimensão.

Para reordenar categorias

Para reordenar categorias manualmente:

1. Selecione (clique) uma categoria na lista.
2. Clique na seta para cima ou para baixo para mover a categoria para cima ou para baixo na lista.

ou

3. Clique na coluna Valor(es) para a categoria e arraste e solte-a em uma posição diferente.

Para excluir categorias

1. Selecione (clique) uma categoria na lista.
2. Clique na seta próxima à lista Exclusão.

ou

3. Clique na coluna Valor(es) para a categoria e arraste e solte-a em qualquer lugar fora da lista.

Se você excluir quaisquer categorias, todas as categorias sem rótulos de valor definido também serão excluídas.

Para ordenar categorias

É possível ordenar categorias por valor dos dados, rótulo de valor, contagem de célula ou estatística de sumarização, em ordem crescente ou decrescente.

1. No grupo Ordenar categorias, clique na lista suspensa Por e selecione o critério de ordenação que você deseja usar: valor, rótulo, contagem ou estatística de sumarização (como média, mediana ou modo). As estatísticas básicas disponíveis para ordenação dependem das estatísticas básicas selecionadas para exibição na tabela.
2. Clique na lista suspensa Ordem para selecionar a ordenação (crescente ou decrescente).

A ordenação de categorias não estará disponível se você tiver excluído quaisquer categorias.

Subtotais

1. Selecione (clique) na categoria na lista que é a última categoria no intervalo de categorias que você deseja incluir no subtotal.
2. Clique em **Incluir subtotal...**
3. Na caixa de diálogo Definir subtotal, opcionalmente, modifique o texto do rótulo do subtotal.
4. Para mostrar apenas um subtotal e suprimir a exibição das categorias que definem o subtotal, selecione **Ocultar categorias subtotalizadas da tabela**.
5. Clique em **Continuar** para incluir o subtotal.

Totais

1. Clique na caixa de seleção **Total**. Também é possível modificar o texto do rótulo.

Se a variável selecionada estiver aninhada dentro de outra variável, os totais serão inseridos para cada subtabela.

Posição de exibição para totais e subtotais

Os totais e subtotais podem ser exibidos acima ou abaixo das categorias incluídas em cada total.

- Se **Abaixo** estiver selecionado no grupo Aparecimento de totais e subtotais, os totais aparecerão acima de cada subtabela e todas as categorias acima e incluindo a categoria selecionada (mas abaixo de qualquer subtotal anterior) serão incluídas em cada subtotal.
- Se **Acima** estiver selecionado no grupo Aparecimento de totais e subtotais, os totais aparecerão abaixo de cada subtabela e todas as categorias abaixo e incluindo a categoria selecionada (mas acima de qualquer subtotal anterior) serão incluídas em cada subtotal.

Importante: Você deve selecionar a posição de exibição de subtotais antes de definir quaisquer subtotais. A mudança da posição de exibição afeta todos os subtotais (não apenas o subtotal selecionado atualmente), e também *muda as categorias incluídas nos subtotais*.

Categorias calculadas

É possível exibir categorias calculadas a partir de estatísticas básicas, totais, subtotais e/ou constantes. Consulte o tópico “Categorias calculadas” para obter mais informações

estatísticas básicas de totais e subtotais customizados

É possível exibir estatísticas diferentes de "totais" nas áreas Totais e Subtotais da tabela usando a caixa de diálogo estatísticas básicas. Consulte o tópico “estatísticas básicas para variáveis categóricas” na página 7 para obter mais informações

Nota: Se você selecionar múltiplas estatísticas de totais customizados que também estão no corpo da tabela e ocultar os rótulos de estatísticas, os totais serão reordenados na mesma ordem que no corpo da tabela — e como os rótulos não são exibidos, não é possível saber o que cada estatística de total realmente representa. Em geral, selecionar múltiplas estatísticas e ocultar os rótulos de estatísticas provavelmente não é uma boa ideia.

Totais, subtotais e categorias excluídas

Os casos de categorias excluídas não estão incluídos no cálculo de totais.

Valores omissos, Categorias vazias e Valores sem rótulos de valor

Valores omissos. Isso controla a exibição de valores **omissos de usuário** ou valores definidos como valores omissos de contenção (por exemplo, um código de 99 para representar "não aplicável" para gravidez em homens). Por padrão, os valores omissos de usuário são excluídos. Selecione (marque) essa opção para incluir categorias com usuário desconhecido nas tabelas. Embora a variável possa conter mais de uma categoria de valor omissos, a visualização da tabela na tela exibirá apenas uma categoria de valor omissos genérico. Todas as categorias com usuário desconhecido definidas serão incluídas na tabela. Os **valores omissos do sistema** (células vazias para variáveis numéricas no Editor de dados) são sempre excluídos.

Categorias vazias. Categorias vazias são categorias com rótulos de valor definido, mas sem casos nessa categoria para uma determinada tabela ou subtabela. Por padrão, as categorias vazias são incluídas em tabelas. Cancele a seleção dessa opção para excluir categorias omissas da tabela.

Outros valores localizados quando os dados são varridos. Por padrão, os valores de categoria no arquivo de dados que não possuem rótulos de valor definido são automaticamente incluídos em tabelas. Cancele a seleção dessa opção para excluir da tabela valores sem rótulos de valor definido. Se você excluir categorias com rótulos de valor definido, as categorias sem rótulos de valor definido também serão excluídas.

Categorias calculadas

Além de exibir os resultados agregados de estatísticas básicas, uma tabela pode exibir uma ou mais categorias calculadas desses resultados agregados, de valores constantes, de subtotais e totais, ou uma combinação deles. Os resultados são conhecidos como categorias calculadas ou pós-cálculos. Uma categoria calculada age como uma categoria em uma única variável com as seguintes semelhanças e diferenças:

- Uma categoria calculada é posicionada como as outras categorias.
- Uma categoria calculada opera nas mesmas estatísticas que as outras categorias.
- As categorias calculadas não afetam subtotais, totais ou testes de significância.
- Por padrão, os valores de categorias calculadas usam a mesma formatação para estatísticas básicas que as outras categorias. É possível substituir o formato ao definir a categoria calculada.

Como as categorias calculadas podem ser usadas para o total de resultados agregados, elas podem ser semelhantes aos subtotais. No entanto, as categorias calculadas possuem as seguintes vantagens sobre os subtotais:

- As categorias calculadas podem ser calculadas a partir dos resultados de outros subtotais.
- As categorias calculadas podem se sobrepor, operando nas mesmas (ou em algumas das mesmas) categorias.
- As categorias calculadas não precisam incluir valores de todas as outras categorias acima ou abaixo da categoria calculada. Ou seja, as categorias calculadas não são exaustivas.
- As categorias calculadas podem incluir valores de categorias que não são adjacentes.

Diferentemente de totais e subtotais, as categorias calculadas são calculadas a partir dos dados agregados em vez dos dados originais. Portanto, os valores de categorias calculadas podem não corresponder aos resultados de totais e subtotais. Além disso, como você tem a opção de ocultar categorias de origem ao definir a categoria calculada, pode ser difícil interceptar subtotais na tabela resultante. Se você usar categorias calculadas, é recomendável especificar rótulos customizados para subtotais.

Para definir uma categoria calculada

As categorias calculadas são incluídas na caixa de diálogo Categorias e totais. Para obter informações sobre como acessar essa caixa de diálogo, consulte o tópico “Categorias e totais” na página 12.

1. Na caixa de diálogo Categorias e totais, clique em **Incluir categoria...**
2. Em **Rótulo para categoria calculada**, especifique um rótulo para a categoria calculada. É possível arrastar categorias da lista Categorias para incluir rótulos para essas categorias.
3. Construa uma expressão selecionando categorias e/ou totais e subtotais e usando operadores para definir as categorias calculadas. Também é possível digitar valores constantes (por exemplo, 500) para incluir na expressão.
4. Para mostrar somente uma categoria calculada e suprimir a exibição das categorias que definem a categoria calculada, selecione **Ocultar categorias usadas na expressão da tabela**.
5. Clique na guia **Formatos de exibição** para mudar o formato de exibição e o número de casas decimais para a categoria calculada. Consulte o tópico “Formatos de exibição para categorias calculadas” para obter mais informações
6. Clique em **Continuar** para incluir a categoria calculada.

Formatos de exibição para categorias calculadas: Por padrão, uma categoria calculada usa o mesmo formato de exibição e número de casas decimais que as outras categorias na variável. É possível substituí-las na guia Formatos de exibição na caixa de diálogo Categoria calculada. A guia Formatos de exibição lista as estatísticas básicas atuais nas quais a categoria calculada opera, além dos formatos de exibição e do número de casas decimais para essas estatísticas.

Para cada estatística de sumarização, é possível:

1. Selecionar um formato de exibição da lista suspensa Formato para a estatística de sumarização. Para obter uma lista completa de formatos de exibição, consulte o tópico “Formatos de exibição de estatísticas básicas” na página 11.
2. Insira o número de casas decimais a serem exibidas na célula Casas decimais para a estatística de sumarização selecionada.

Tabelas de variáveis com categorias compartilhadas (Tabelas Comperimeter)

As pesquisas de opinião geralmente contêm muitas perguntas com um conjunto comum de respostas possíveis. É possível usar o empilhamento para exibir essas variáveis relacionadas na mesma tabela, e exibir as categorias de resposta compartilhadas nas colunas da tabela.

Para criar uma tabela para variáveis múltiplas com categorias compartilhadas

Tabela 6. Variáveis empilhadas com categorias de resposta compartilhadas em colunas

Variáveis	Categoria 1	Categoria 2	Categoria 3
Variável 1	12	34	56
Variável 2	56	12	34
Variável 3	34	56	12

Consulte o tópico “Tabelas para variáveis com categorias compartilhadas” na página 41 para obter mais informações

Customizando o construtor de tabela

Diferentemente das caixas de diálogo padrão, é possível mudar o tamanho do construtor de tabela da mesma forma que é possível mudar o tamanho de qualquer janela padrão:

1. Clique e arraste a parte superior, a parte inferior, qualquer lado ou qualquer canto do construtor de tabela para diminuir ou aumentar seu tamanho.
Na guia Tabela, também é possível mudar o tamanho da lista de variáveis, da lista Categorias e da área de janela de tela.
2. Clique e arraste a barra horizontal entre a lista de variáveis e a lista Categorias para aumentar ou diminuir as listas. Movê-la para baixo aumenta a lista de variáveis e diminui a lista Categorias. Movê-la para cima tem o efeito contrário.
3. Clique e arraste a barra vertical entre a lista de variáveis e a lista Categorias na área de janela de tela para tornar as listas mais largas ou mais estreitas. A tela é redimensionada automaticamente para se ajustar ao espaço restante.

Tabelas customizadas: Guia Opções

Aparência da célula de dados

Controla o que é exibido em células vazias e as células para as quais as estatísticas não podem ser calculadas.

Células vazias

Para células de tabela que não contêm casos (contagem de célula de 0), é possível selecionar uma de três opções de exibição: zero, em branco ou um valor de texto que for especificado. O valor de texto pode ter até 255 caracteres.

Estatísticas que não podem ser calculadas

Texto que é exibido se uma estatística não puder ser calculada (por exemplo, a média para uma categoria sem nenhum caso). O valor de texto pode ter até 255 caracteres. O valor padrão é um ponto (.).

Largura para a coluna de dados

Controla as larguras mínima e máxima da coluna para colunas de dados. Essa configuração não afeta as larguras de colunas para rótulos da linha.

Configurações do TableLook

Usa a especificação de largura da coluna de dados do TableLook padrão atual. É possível criar seu próprio TableLook padrão customizado para ser usado quando novas tabelas forem criadas, e é possível controlar larguras da coluna de rótulo da linha e da coluna de dados com um TableLook.

Personalizado

Substitui as configurações de TableLook padrão para a largura da coluna de dados. Especifique as larguras mínima e máxima da coluna de dados para a tabela e a unidade de medida: pontos, polegadas ou centímetros.

Valores omissos para variáveis de escala

Para tabelas com duas ou mais variáveis de escala, controla o tratamento de dados omissos para estatísticas de variável de escala.

Maximizar o uso dos dados disponíveis (exclusão por variável)

Todos os casos com valores válidos para cada variável de escala são incluídos nas estatísticas básicas para essa variável de escala.

Usar base de dados consistente através de variáveis de escala (exclusão de listwise)

Os casos com valores omissos para qualquer variável de escala na tabela são excluídos das estatísticas básicas para todas as variáveis de escala na tabela.

Base Efetiva

Se você tiver uma variável que representa ponderações de ajustamento em vez de ponderações de frequência, será possível usar essa variável como uma variável de ponderação de base efetiva. O conceito de peso de tamanho da base efetiva ou da amostra efetiva é baseado em métodos para a análise de dados a partir de amostras complexas. Uma ponderação de base efetiva permite tratamento aproximado de inferência estatística em análise que envolve ajustamentos ad hoc para dados a partir de planos de amostragem aleatória simples, usando ponderações de ajustamento.

- A ponderação de base efetiva afeta valores de estatísticas básicas ponderados e testes de significância de médias de coluna e de proporções da coluna.
- Se a ponderação estiver ativada para o conjunto de dados, a variável de ponderação do conjunto de dados será ignorada e os resultados serão ponderados pela variável de ponderação de base efetiva.
- A variável de ponderação de base efetiva deve ser numérica.
- Os casos com valores de ponderação negativos, um valor de ponderação de 0 ou valores de ponderação omissos são excluídos de todos os resultados.

Contar respostas duplicadas para conjuntos de múltiplas categorias

Uma resposta duplicada é a mesma resposta para duas ou mais variáveis no conjunto de categorias múltiplas. Por padrão, as respostas duplicadas não são contadas.

Ocultar contagens pequenas

É possível optar por ocultar contagens que menores que um número inteiro especificado.

- Os valores ocultos são exibidos como $<N$, em que N é o número inteiro especificado.
- O número inteiro especificado deve ser maior ou igual a 2.
- Se a ponderação estiver ativada para o conjunto de dados ou uma variável de ponderação de base efetiva for especificada, o valor ponderado será usado.

Ponderações e Arredondamentos

- Se você usar **Dados > Casos de Ponderação** para ponderar casos, ponderações não inteiras são arredondadas na célula ou no nível de categoria para testes de significância, intervalos de confiança e erros padrão.
- Se você selecionar **Usar variável de ponderação de base efetiva**, valores de ponderação não inteiros na variável de ponderação selecionada não serão arredondados.
- Se ambos forem especificados, a variável de ponderação de base efetiva será usada e ponderações não inteiras não serão arredondadas.

Tabelas customizadas: Guia Títulos

A guia Títulos controla a exibição de títulos, legendas e rótulos de canto.

Título. Texto que é exibido acima da tabela.

Legenda. Texto que é exibido abaixo da tabela e acima de quaisquer notas de rodapé.

Canto. Texto que é exibido no canto superior esquerdo da tabela. O texto de canto será exibido apenas se a tabela contiver variáveis de linha e se a propriedade de rótulo de dimensão da linha da tabela dinâmica estiver configurada como **Aninhado**. Essa *não* é a configuração de TableLook padrão.

É possível incluir os seguintes valores gerados automaticamente no título da tabela, na legenda ou no rótulo de canto:

Data. Ano, mês e dia atuais exibidos em um formato baseado em suas configurações atuais de Opções regionais do Windows.

Horário. Hora, minuto e segundo atuais exibidos em um formato baseado nas configurações atuais de Opções regionais do Windows.

Expressão de tabela. Variáveis usadas na tabela e como elas serão usadas na tabela. Se uma variável tiver um rótulo de variáveis definido, o rótulo será exibido. Na tabela gerada, os símbolos a seguir indicam como as variáveis são usadas na tabela:

- + indica variáveis empilhadas.
- > indica aninhamento.
- **BY** indica tabulação cruzada ou camadas.

Tabelas customizadas: Guia Estatísticas de teste

A guia Estatísticas de teste fornece testes de significância para tabelas customizadas.

Esses testes não estão disponíveis para tabelas nas quais os rótulos de categoria são movidos para fora da dimensão de tabela padrão ou para categorias calculadas.

Testes de médias de coluna e de proporções da coluna

Os testes de médias de coluna estão disponíveis para variáveis de escala. Os testes de proporções da coluna estão disponíveis para variáveis categóricas.

Comparar médias de colunas

Testes de pares da igualdade de médias de coluna. A tabela deve ter uma variável categórica nas colunas e uma variável de escala como o nível mais interno das linhas. A tabela deve incluir a média como uma estatística de sumarização.

Para variáveis categóricas ordinárias, a variância pode ser estimada a partir de todas as categorias ou apenas das categorias que são comparadas. Para variáveis de múltiplas respostas, a variância para o teste de médias é sempre baseada apenas nas categorias que são comparadas.

Comparar proporções de coluna

Testes de pares da igualdade de proporções da coluna. A tabela deve ter pelo menos uma variável categórica nas colunas e linhas. A tabela deve incluir contagens ou porcentagens da coluna.

Identificar diferenças significativas

Para testes de médias de coluna e de proporções da coluna, é possível exibir resultados significativos em uma tabela separada ou na tabela principal.

Em uma tabela separada

Os resultados de testes de significância são exibidos em uma tabela separada. Se dois valores forem significativamente diferentes, a célula correspondente ao valor maior exibirá uma chave que identifica a coluna com o valor menor.

Exibir valores de significância

Os valores de significância são exibidos entre parênteses após cada valor da chave na célula. Esta opção está disponível apenas quando os resultados significativos são exibidos em uma tabela separada.

Na tabela principal

Os resultados do teste de significância são exibidos na tabela principal. Cada categoria de coluna na tabela é identificada com uma chave alfabética. Para cada par significativo, a chave da categoria com a menor média ou proporção da coluna aparece na categoria com a maior média ou proporção da coluna.

- Quando você passar o mouse sobre uma chave na célula de rótulo da coluna em uma tabela dinâmica, todas as células na tabela com aquela chave de significância serão destacadas. Para uma tabela com variáveis múltiplas na dimensão da coluna, somente células naquela subtabela serão destacadas.
- Para selecionar todas as células em uma tabela (ou subtabela) que têm a mesma chave de significância, clique com o botão direito na célula de rótulo da coluna e escolha **Selecionar > Selecionar todas as células com essa chave de significância**.

Usar subscritos de estilo APA

Identifique diferenças significativas com formatação de estilo APA que usa letras subscritas. Se dois valores forem significativamente diferentes, esses valores exibirão diferentes letras subscritas. Esses subscritos não são notas de rodapé. Quando essa opção estiver em vigor, o estilo de nota de rodapé definido no TableLook atual será substituído e as notas de rodapé serão exibidas como números sobrescritos. Para selecionar todas as células na mesma linha com a mesma chave de significância, clique com o botão direito em uma célula que tem uma chave de significância e escolha **Selecionar células com significância semelhante**

Níveis de significância

O nível de significância para testes de médias de coluna e de proporções da coluna.

- O valor deve ser maior que 0 e menor que 1.
- Se você especificar dois níveis de significância, letras maiúsculas serão usadas para identificar valores de significância menores ou iguais ao menor nível. Letras minúsculas são usadas para identificar valores de significância menores ou iguais ao maior nível.
- Se você selecionar **Usar subscritos de estilo de APA**, o segundo valor será ignorado.

Ajustar valores p para comparações múltiplas

A correção de **Bonferroni** é ajustada para a taxa de erro da família (FWER). O método de **Benjamini-Hochberg** é um ajustamento de taxa de descoberta falsa (FDR). Esse método é menos conservador que a correção de Bonferroni.

Testes de independência (qui-quadrado)

Teste qui-quadrado de independência para tabelas nas quais existe pelo menos uma variável de categoria nas linhas e colunas.

Usar subtotaís em vez de categorias subtotalizadas

Cada subtotal substitui suas categorias para o teste de significância. Caso contrário, apenas os subtotaís para os quais os subtotaís das categorias estão ocultos substituirão suas categorias para teste.

Incluir variáveis de múltiplas respostas nos testes

As categorias de conjuntos de múltiplas respostas são incluídas em testes de significância. Caso contrário, os conjuntos de múltiplas respostas não são incluídos em testes de significância.

Tabelas simples para variáveis categóricas












Tabelas simples para variáveis categóricas

A maioria das tabelas que você deseja criar provavelmente incluirão pelo menos uma **variável categórica**. Uma variável categórica é uma com um número limitado de valores ou categorias distintas (por exemplo, gênero ou religião). As variáveis categóricas podem ser **nominais** ou **ordinais**.

- *Nominal*. Uma variável pode ser tratada como nominal quando seus valores representem categorias sem ranqueamento intrínseco (por exemplo, o departamento da empresa na qual um funcionário trabalha). Exemplos de variáveis nominais incluem região, código de endereçamento postal e filiação religiosa.
- *Ordinal*. Uma variável pode ser tratada como ordinal quando seus valores representem categorias com algum ranqueamento intrínseco (por exemplo, níveis de satisfação de serviço de muito insatisfeito para muito satisfeito). Exemplos de variáveis ordinais incluem escores de atitude que representam o grau de satisfação ou de confiança e os escores de classificação de preferência.

Um ícone próximo a cada variável na lista de variáveis identifica o tipo de variável.

Tabela 7. Ícones do nível de medição

	Numérico	Sequência de caracteres	Data	Horário
Escala (Contínua)		n/a		
Ordinal				
Nominal				

As Tabelas customizadas são otimizadas para uso com variáveis categóricas que possuem **rótulos de valor** definidos. Consulte o tópico “Construindo tabelas” na página 1 para obter mais informações

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *survey_sample.sav*. Consulte o tópico arquivos de dados para obter mais informações

Todos os exemplos fornecidos aqui exibem rótulos de variáveis em caixas de diálogo, ordenados em ordem alfabética. As propriedades de exibição da lista de variáveis são configuradas na guia Geral na caixa de diálogo Opções (menu Editar, Opções).

Uma variável categórica única

Embora uma tabela de uma única variável categórica possa ser uma das tabelas mais simples que pode ser criada, geralmente ela pode ser tudo o que você quer ou precisa.

1. Nos menus, escolha:

Analisar > Tabelas > Tabelas Customizadas...

2. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.

Uma visualização da tabela é exibida na área de janela de tela. A visualização não exibe valores de dados reais; ela exibe somente itens temporários nos quais os dados serão exibidos.

3. Clique em **OK** para criar a tabela.

A tabela é exibida na janela Visualizador.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Figura 2. Variável categórica única em linhas

Nesta tabela simples, o título da coluna *Contagem* não é realmente necessário, e é possível criar a tabela sem esse título da coluna.

- Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
- No grupo estatísticas básicas, selecione (clique) **Ocultar** para Posição.
- Clique em **OK** para criar a tabela.

Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Figura 3. Variável categórica única sem rótulo da coluna de estatísticas básicas

Porcentagens

Além das contagens, também é possível exibir porcentagens. Para uma tabela simples de uma única variável categórica, se a variável for exibida em linhas, provavelmente você deseja examinar porcentagens da coluna. Por outro lado, para uma variável exibida em colunas, provavelmente você deseja examinar porcentagens de linhas.

- Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
- No grupo estatísticas básicas, cancele a seleção de **Ocultar** para Posição. Como essa tabela terá duas colunas, você deseja exibir os rótulos da coluna, portanto, sabe o que cada coluna representa.
- Clique com o botão direito em *Categoria de idade* na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
- Na caixa de diálogo estatísticas básicas, selecione **% de N da coluna** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
- Na célula Rótulo na lista Exibição, exclua o rótulo padrão e digite Percentual.
- Clique em **Aplicar à seleção** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%

Figura 4. Contagens e porcentagens da coluna

Totais

Os totais não são automaticamente incluídos nas tabelas customizadas, mas é fácil incluir totais em uma tabela.

- Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).

2. Clique com o botão direito em *Categoria de idade* na área de janela de tela e selecione **Categorias e totais** no menu pop-up.
3. Selecione (clique) **Total** na caixa de diálogo Categorias e totais.
4. Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	Total	2828	100.0%

Figura 5. Contagens, porcentagens da coluna e totais

Consulte o tópico “Totais e subtotais para variáveis categóricas” na página 32 para obter mais informações

Tabulação cruzada

Tabulação cruzada é uma técnica básica para examinar o relacionamento entre duas variáveis categóricas. Por exemplo, usando *Categoria de idade* como uma variável de linha e *Gênero* como uma variável da coluna, é possível criar uma tabulação cruzada bidimensional que mostra o número de homens e mulheres em cada categoria de idade.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em **Reconfigurar** para excluir todas as seleções anteriores no construtor de tabela.
3. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.
4. Arraste e solte *Gênero* da lista de variáveis para a área Colunas na área de janela de tela. (Pode ser necessário rolar para baixo pela lista de variáveis para localizar essa variável.)
5. Clique em **OK** para criar a tabela.

		Gender	
		Male	Female
		Count	Count
Age category	Less than 25	108	134
	25 to 34	276	351
	35 to 44	309	370
	45 to 54	221	260
	55 to 64	136	184
	65 or older	178	301

Figura 6. Tabulação cruzada de *Categoria de idade* e *Gênero*

Porcentagens em tabulações cruzadas

Em uma tabulação cruzada bidimensional, as porcentagens de linhas e da coluna podem fornecer informações úteis.

1. Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
2. Clique com o botão direito em *Gênero* na área de janela de tela.

É possível notar que a opção **estatísticas básicas** está desativada no menu pop-up. Isso ocorre porque é possível selecionar estatísticas básicas apenas para a variável mais interna na dimensão de origem estatística. A dimensão de origem estatística padrão (linha ou coluna) para variáveis categóricas é baseada na ordem em que você arrasta e solta variáveis na área de janela de tela. Nesse exemplo, arrastamos *Categoria de idade* para a dimensão de linhas primeiro -- e como não há outras variáveis na

dimensão de linhas, *Categoria de idade* é a variável de origem estatística. É possível mudar a dimensão de origem estatística, mas nesse exemplo, não é necessário fazer isso. Consulte o tópico “Estatísticas básicas” na página 6 para obter mais informações

3. Clique com o botão direito em *Categoria de idade* na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
4. Na caixa de diálogo estatísticas básicas, selecione **% de N da coluna** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
5. Selecione **% de N da linha** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
6. Clique em **Aplicar à seleção** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Gender					
		Male			Female		
		Count	Column N %	Row N %	Count	Column N %	Row N %
Age category	Less than 25	108	8.8%	44.6%	134	8.4%	55.4%
	25 to 34	276	22.5%	44.0%	351	21.9%	56.0%
	35 to 44	309	25.2%	45.5%	370	23.1%	54.5%
	45 to 54	221	18.0%	45.9%	260	16.3%	54.1%
	55 to 64	136	11.1%	42.5%	184	11.5%	57.5%
	65 or older	178	14.5%	37.2%	301	18.8%	62.8%

Figura 7. Tabulação cruzada com porcentagens de linhas e da coluna

Controlando o formato de exibição

É possível controlar o formato de exibição, incluindo o número de decimais exibidos nas estatísticas básicas. Por exemplo, por padrão, as porcentagens são exibidas como um decimal e um sinal de porcentagem. Mas e se você desejar que os valores da célula mostrem dois decimais e nenhum sinal de porcentagem?

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique com o botão direito em *Categoria de idade* na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
3. Para as duas estatísticas básicas selecionadas (**% de N da coluna** e **% de N da linha**), selecione **nnnn.n** da lista suspensa Formato e digite 2 na célula Casas Decimais para ambas.
4. Clique em **OK** para criar a tabela.

		Gender					
		Male			Female		
		Count	Column N %	Row N %	Count	Column N %	Row N %
Age category	Less than 25	108	8.79	44.63	134	8.38	55.37
	25 to 34	276	22.48	44.02	351	21.94	55.98
	35 to 44	309	25.16	45.51	370	23.13	54.49
	45 to 54	221	18.00	45.95	260	16.25	54.05
	55 to 64	136	11.07	42.50	184	11.50	57.50
	65 or older	178	14.50	37.16	301	18.81	62.84

Figura 8. Exibição de célula formatada para porcentagens de linhas e da coluna

Totais marginais

É bastante comum em tabulações cruzadas a exibição de **totais marginais** -- totais para cada linha e coluna. Como eles não são incluídos em Tabelas customizadas por padrão, é necessário incluí-los explicitamente em suas tabelas.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em **Reconfigurar** para excluir todas as seleções anteriores no construtor de tabela.
3. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.

4. Arraste e solte *Gênero* da lista de variáveis para a área Colunas na área de janela de tela. (Pode ser necessário rolar para baixo pela lista de variáveis para localizar essa variável.)
5. Clique com o botão direito em *Categoria de idade* na área de janela de tela e selecione **Categorias e totais** no menu pop-up.
6. Selecione (clique) **Total** na caixa de diálogo Categorias e totais e, em seguida, clique em **Aplicar**.
7. Clique com o botão direito em *Gênero* na área de janela de tela e selecione **Categorias e totais** no menu pop-up.
8. Selecione (clique) **Total** na caixa de diálogo Categorias e totais e, em seguida, clique em **Aplicar**.
9. No grupo estatísticas básicas, selecione (clique) **Ocultar** para Posição. (Como você está exibindo apenas contagens, não precisa identificar a "estatística" exibida nas células de dados da tabela.)
10. Clique em **OK** para criar a tabela.

		Gender		
		Male	Female	Total
Age category	Less than 25	108	134	242
	25 to 34	276	351	627
	35 to 44	309	370	679
	45 to 54	221	260	481
	55 to 64	136	184	320
	65 or older	178	301	479
	Total	1228	1600	2828

Figura 9. Tabulação cruzada com totais marginais

Ordenando e excluindo categorias

Por padrão, as categorias são exibidas na ordem crescente dos valores de dados que os rótulos de valor de categoria representam. Por exemplo, embora os rótulos de valor de *Menor que 25, 25, 34, 35, 44,...*, etc., sejam exibidos para categorias de idade, os valores de dados subjacentes reais são 1, 2, 3, ..., etc., e são esses valores de dados subjacentes que controlam a ordem de exibição padrão das categorias.

É possível mudar facilmente a ordem das categorias e também excluir categorias que você não deseja que sejam exibidas na tabela.

Ordenando categorias

É possível reorganizar manualmente categorias ou ordenar categorias em ordem crescente ou decrescente de:

- Valores de dados.
- Rótulos de valor.
- Contagens de célula.
- estatísticas básicas. As estatísticas básicas disponíveis para ordenação dependem das estatísticas básicas selecionadas para exibição na tabela.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Se *Categoria de idade* ainda não estiver exibida na área Linhas na área de janela de tela, arraste e solte-a lá.
3. Clique com o botão direito em *Categoria de idade* na área de janela de tela e selecione **Categorias e totais** no menu pop-up.

Os valores de dados e os rótulos de valor associados são exibidos na ordenação atual que, nesse caso, ainda é ordem crescente de valores de dados.

4. No grupo Ordenar categorias, selecione **Decrescente** na lista suspensa Ordem.
A ordenação agora é invertida.
5. Selecione **Rótulos** na lista suspensa Por.

As categorias agora são ordenadas em ordem alfabética decrescente dos rótulos de valor.

Observe que a categoria rotulada *Menor que 25* está na parte superior da lista. Em ordenação alfabética, as letras vêm após os números. Como esse é o único rótulo que começa com uma letra e a lista é ordenada em ordem decrescente (inversa), essa categoria é ordenada na parte superior da lista.

Se desejar que uma determinada categoria apareça em um local diferente na lista, é possível movê-la facilmente.

6. Clique na categoria rotulada *Menor que 25* na lista Rótulo.
7. Clique na seta para baixo à direita da lista. A categoria é movida uma linha para baixo na lista.
8. Continue clicando na seta para baixo até que a categoria esteja na parte inferior da lista.

Excluindo categorias

Se houver algumas categorias que você não deseja que apareçam na tabela, é possível excluí-las.

1. Clique na categoria rotulada *Menor que 25* na lista Rótulo.
2. Clique na tecla de seta à esquerda da lista Exclusão.
3. Clique na categoria rotulada *65 ou mais* na lista Rótulo.
4. Clique na tecla de seta à esquerda da lista Exclusão novamente.
As duas categorias são movidas da lista Exibição para a lista Exclusão. Se mudar de ideia, é possível movê-las facilmente de volta para a lista Exibição.
5. Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Gender		
		Male	Female	Total
Age category	55 to 64	136	184	320
	45 to 54	221	260	481
	35 to 44	309	370	679
	25 to 34	276	351	627
	Total	942	1165	2107

Figura 10. Tabela ordenada por rótulo de valor decrescente, algumas categorias excluídas

Observe que os totais são inferiores do que eram antes da exclusão das duas categorias. Isso ocorre porque os totais são baseados nas categorias incluídas na tabela. Todas as categorias excluídas são excluídas do cálculo total. Consulte o tópico “Totais e subtotais para variáveis categóricas” na página 32 para obter mais informações

Empilhamento, aninhamento e estratos com variáveis categóricas

Empilhamento, aninhamento e estratos são todos métodos para exibir variáveis múltiplas na mesma tabela. Esse capítulo focaliza o uso dessas técnicas com variáveis categóricas, embora elas também possam ser usadas com variáveis de escala.

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *survey_sample.sav*. Consulte o tópico arquivos de dados para obter mais informações

Todos os exemplos fornecidos aqui exibem rótulos de variáveis em caixas de diálogo, ordenados em ordem alfabética. As propriedades de exibição da lista de variáveis são configuradas na guia Geral na caixa de diálogo Opções (menu Editar, Opções).

Empilhando variáveis categóricas

O empilhamento pode ser considerado como pegar tabelas separadas e colá-las juntas na mesma exibição. Por exemplo, é possível exibir informações sobre *Gênero* e *Categoria de idade* em seções separadas da mesma tabela.

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. No construtor de tabela, arraste e solte *Gênero* da lista de variáveis para a área Linhas na área de janela de tela.
3. Arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas, abaixo de *Gênero*.
As duas variáveis agora estão empilhadas na dimensão da linha.
4. Clique em **OK** para criar a tabela.

		Count
Gender	Male	1232
	Female	1600
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Figura 11. Tabela de variáveis categóricas empilhadas em linhas

Também é possível empilhar variáveis em colunas de uma forma semelhante.

Empilhamento com tabulação cruzada

Uma tabela empilhada pode incluir outras variáveis em outras dimensões. Por exemplo, é possível fazer a tabulação cruzada de duas variáveis empilhadas nas linhas com uma terceira variável exibida na dimensão da coluna.

1. Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
2. Se *Categoria de idade* e *Gênero* ainda não estiverem empilhados nas linhas, siga as instruções acima para empilhá-los.
3. Arraste e solte *Obter notícias da Internet* da lista de variáveis para a área Colunas na área de janela de tela.
4. Clique em **OK** para criar a tabela.

		Get news from internet	
		No	Yes
		Count	Count
Gender	Male	873	359
	Female	1092	508
Age category	Less than 25	146	96
	25 to 34	368	259
	35 to 44	435	244
	45 to 54	346	135
	55 to 64	252	68
	65 or older	416	63

Figura 12. Duas variáveis de linha empilhadas com tabulação cruzada com uma variável da coluna

Nota: Existem diversas variáveis com rótulos que começam com *Obter notícias de ...*, portanto, pode ser difícil distinguir entre elas na lista de variáveis (já que os rótulos podem ser muito grandes para serem exibidos completamente na lista de variáveis). Existem duas maneiras de ver o rótulo de variáveis inteiro:

- Posicione o ponteiro do mouse sobre uma variável na lista para exibir o rótulo inteiro em uma Dica de ferramenta pop-up.
- Clique e arraste a barra vertical que separa as listas de variáveis e de Categorias da área de janela de tela para tornar as listas mais largas.

Aninhando variáveis categóricas

O aninhamento, assim como a tabulação cruzada, pode mostrar o relacionamento entre duas variáveis categóricas, exceto que uma variável está aninhada na outra na mesma dimensão. Por exemplo, é possível aninhar *Gênero* em *Categoria de idade* na dimensão da linha, mostrando o número de homens e mulheres em cada categoria de idade.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em **Reconfigurar** para excluir todas as seleções anteriores no construtor de tabela.
3. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.
4. Arraste e solte *Gênero* da lista de variáveis à direita de *Categoria de idade* na área Linhas.

A visualização na área de janela de tela agora mostra que a tabela aninhada conterá uma única coluna de contagens, com cada célula contendo o número de homens ou mulheres em cada categoria de idade.

É possível notar que o rótulo de variáveis *Gênero* é exibido repetidamente, uma vez para cada categoria de idade. É possível minimizar esse tipo de repetição colocando a variável com as menores quantidades de categorias no nível mais externo do aninhamento.

5. Clique no rótulo de variáveis *Gênero* na área de janela de tela.
6. Arraste e solte a variável o máximo possível à esquerda na área Linhas.

Agora, em vez de *Gênero* estar sendo repetido seis vezes, *Categoria de idade* é repetida duas vezes. Essa é uma tabela menos desordenada que produzirá basicamente os mesmos resultados.

7. Clique em **OK** para criar a tabela.

				Count
Gender	Male	Age category	Less than 25	108
			25 to 34	276
			35 to 44	309
			45 to 54	221
			55 to 64	136
			65 or older	178
	Female	Age category	Less than 25	134
			25 to 34	351
			35 to 44	370
			45 to 54	260
			55 to 64	184
			65 or older	301

Figura 13. Tabela de Categoria de idade aninhada em Gênero

Nota: As Tabelas customizadas não consideram o processamento de arquivo dividido em estratos. Para obter o mesmo resultado como arquivos divididos em estratos, coloque as variáveis de arquivo dividido nas estratos de aninhamento mais externas da tabela.

Suprimindo rótulos de variáveis

Outra solução para rótulos de variáveis redundantes em tabelas aninhadas é simplesmente suprimir a exibição de nomes ou rótulos de variáveis. Como os rótulos de valor para *Gênero* e *Categoria de idade* provavelmente são suficientemente descritivos sem os rótulos de variáveis, podemos eliminar os rótulos para ambas as variáveis.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).

- Clique com o botão direito em *Categoria de idade* na área de janela de tela e cancele a seleção de **Mostrar rótulo de variáveis** no menu pop-up.
- Faça o mesmo para *Gênero*.
Os rótulos de variáveis ainda são exibidos na visualização da tabela, mas eles não serão incluídos na tabela.
- Clique em **OK** para criar a tabela.

		Count
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Figura 14. Tabela aninhada sem rótulos de variáveis

Se desejar que os rótulos de variáveis sejam incluídos com a tabela em algum lugar -- sem exibi-los várias vezes no corpo da tabela -- é possível incluí-los no rótulo do título ou do canto da tabela.

- Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
- Clique na guia **Títulos**.
- Clique em qualquer lugar na caixa de texto Título.
- Clique em **Expressão de tabela**. O texto $\&[Expressão\ de\ tabela]$ é exibido na caixa de texto Título. Isso gerará um título de tabela que inclui os rótulos de variáveis para as variáveis usadas na tabela.
- Clique em **OK** para criar a tabela.

Gender > Age category

		Count
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Figura 15. Rótulos de variáveis no título da tabela

O sinal maior que (>) no título indica que *Categoria de idade* está aninhado em *Gênero*.

Tabulação cruzada aninhada

Uma tabela aninhada pode conter outras variáveis em outras dimensões. Por exemplo, é possível aninhar *Categoria de idade* em *Gênero* nas linhas e fazer a tabulação cruzada das linhas aninhadas em uma terceira variável na dimensão da coluna.

- Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).

2. Se *Categoria de idade* ainda não estiver aninhada dentro de *Sexo* nas linhas, siga as orientações no exemplo anterior para aninhá-las.
3. Arraste e solte *Obter notícias da Internet* da lista de variáveis para a área Colunas na área de janela de tela.

É possível notar que a tabela é muito grande para ser exibida completamente na área de janela de tela. É possível rolar para cima/para baixo ou para a direita/esquerda na área de janela de tela para ver mais da visualização da tabela, ou é possível:

- Clicar em **Compacto** no construtor de tabela para ver uma visualização compacta. Isso exibe apenas os rótulos de variáveis, sem nenhuma informação sobre as categorias ou estatísticas básicas incluídas na tabela.
 - Aumentar o tamanho do construtor de tabela clicando e arrastando qualquer um dos lados ou cantos do construtor de tabela.
4. Clique em **OK** para criar a tabela.

				Get news from internet	
				No	Yes
				Count	Count
Gender	Male	Age category	Less than 25	59	49
			25 to 34	159	117
			35 to 44	217	92
			45 to 54	169	52
			55 to 64	112	24
			65 or older	155	23
	Female	Age category	Less than 25	87	47
			25 to 34	209	142
			35 to 44	218	152
			45 to 54	177	83
			55 to 64	140	44
			65 or older	261	40

Figura 16. Tabulação cruzada aninhada

Trocando linhas e colunas: O que fazer se você gastar muito tempo configurando uma tabela complexa e depois decidir que ela está absolutamente perfeita -- exceto que você deseja alternar a orientação, colocando todas as variáveis de linha nas colunas e vice-versa? Por exemplo, você criou uma tabulação cruzada aninhada com *Categoria de idade* e *Gênero* aninhados nas linhas, mas, em vez disso, agora deseja essas duas variáveis demográficas aninhadas nas colunas.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique com o botão direito em qualquer lugar na área de janela de tela e selecione **Trocar variáveis de linha e de coluna** no menu pop-up.

As variáveis de linha e de coluna agora foram alternadas.

Antes de criar a tabela, vamos fazer algumas modificações para tornar a exibição menos desordenada.

3. Selecione **Ocultar** para suprimir a exibição do rótulo da coluna de estatísticas básicas.
4. Clique com o botão direito em *Gênero* na área de janela de tela e cancele a seleção de **Mostrar rótulo de variáveis**.
5. Agora clique em **OK** para criar a tabela.

		Male						Female					
		Age category						Age category					
		Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older	Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older
Get news from internet	No	59	159	217	169	112	155	87	209	218	177	140	261
	Yes	49	117	92	52	24	23	47	142	152	83	44	40

Figura 17. Tabulação cruzada com variáveis demográficas aninhadas em colunas

Camadas

É possível usar estratos para incluir uma dimensão de espessura em suas tabelas, criando "cubos tridimensionais." As estratos são, de fato, muito semelhantes ao aninhamento ou empilhamento; a principal diferença é que apenas uma categoria de estrato fica visível por vez. Por exemplo, usar *Categoria de idade* como a variável de linha e *Gênero* como uma variável de estrato produz uma tabela na qual as informações para homens e mulheres são exibidas em diferentes estratos da tabela.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em **Reconfigurar** para excluir todas as seleções anteriores no construtor de tabela.
3. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.
4. Clique em **Camadas** na parte superior do construtor de tabela para exibir a lista Camadas.
5. Arraste e solte *Gênero* da lista de variáveis para a lista Camadas.

Nesse ponto, você pode notar que a inclusão de uma variável de estrato não tem nenhum efeito visível na visualização exibida na área de janela de tela. As variáveis de estrato não afetam a visualização na área de janela de tela, a menos que a variável de estrato seja a variável de origem estatística e você mude as estatísticas básicas.

6. Clique em **OK** para criar a tabela.

		Count
Age category	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178

Figura 18. Tabela de estratos simples

À primeira vista, esta tabela não parece diferente de uma tabela simples de uma variável categórica única. A única diferença é a presença do rótulo *Gênero masculino* na parte superior da tabela.

7. Clique duas vezes na tabela na janela Visualizador para ativá-la.
8. Agora é possível ver que o rótulo *Gênero masculino* é realmente uma opção em uma lista suspensa.
9. Clique na seta para baixo na lista suspensa para exibir toda a lista de estratos. Nessa tabela, há apenas uma outra opção na lista.
10. Selecione *Gênero feminino* na lista suspensa.

		Count
Age category	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Figura 19. Tabela de estrato simples com uma estrato diferente exibida

Duas variáveis de estrato categóricas empilhadas

Se você tiver mais de uma variável categórica nas estratos, será possível empilhar ou aninhar as variáveis de estrato. Por padrão, as variáveis de estrato são empilhadas. (Nota: Se você tiver quaisquer variáveis de estrato de escala, as variáveis de estrato somente poderão ser empilhadas.)

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Se você ainda não tem *Categoria de idade* nas linhas e *Sexo* nas camadas, siga as orientações no exemplo anterior para criar uma tabela colocada em camadas.
3. Arraste e solte *Grau mais alto* da lista de variáveis para a lista Camada, abaixo de *Gênero*.

Os dois botões de opções abaixo da lista Camada no grupo Saída de estrato agora estão ativados. A seleção padrão é **Mostrar cada categoria como uma estrato**. Isso é equivalente ao empilhamento.

4. Clique em **OK** para criar a tabela.
5. Clique duas vezes na tabela na janela Visualizador para ativá-la.
6. Clique na seta para baixo na lista suspensa para exibir toda a lista de estratos.

Existem sete estratos na tabela: duas estratos para as duas categorias *Gênero* e cinco estratos para as cinco categorias *Grau mais alto*. Para estratos empilhadas, o número total de estratos é a soma do número de categorias para as variáveis de estrato (incluindo qualquer total ou subtotal de categorias solicitadas para as variáveis de estrato).

Duas variáveis de estrato categóricas aninhadas

O aninhamento de variáveis de estrato categóricas cria uma estrato separada para cada combinação de categorias de variável de estrato.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Se você ainda não fez isso, siga as orientações no exemplo anterior para criar uma tabela de camadas empilhadas.
3. No grupo Saída de estrato, selecione **Mostrar cada combinação de categorias como uma estrato**. Isto é equivalente ao aninhamento.
4. Clique em **OK** para criar a tabela.
5. Clique duas vezes na tabela na janela Visualizador para ativá-la.
6. Clique na seta para baixo na lista suspensa para exibir toda a lista de estratos.

Existem 10 estratos na tabela (é necessário rolar pela lista para ver todas elas), uma para cada combinação de *Gênero* e *Grau mais alto*. Para estratos aninhadas, o número total de estratos é o *produto* do número de categorias para cada variável de estrato (nesse exemplo, $5 \times 2 = 10$).

Imprimindo tabelas em estratos: Por padrão, apenas a estrato visível atualmente é impressa. Para imprimir todas as estratos de uma tabela:

1. Clique duas vezes na tabela na janela Visualizador para ativá-la.
2. Nos menus da janela Visualizador, escolha:

Formatar > Propriedades da Tabela...

3. Clique na guia **Impressão**.
4. Selecione **Imprimir todas as estratos**.

Também é possível salvar essa configuração como parte de um TableLook, incluindo o TableLook padrão.

Totais e subtotais para variáveis categóricas

É possível incluir totais e subtotais em tabelas customizadas. Os totais e subtotais podem ser aplicados a variáveis categóricas em qualquer nível de aninhamento em qualquer dimensão -- linha, coluna ou estrato.

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *survey_sample.sav*. Consulte o tópico arquivos de dados para obter mais informações

Todos os exemplos fornecidos aqui exibem rótulos de variáveis em caixas de diálogo, ordenados em ordem alfabética. As propriedades de exibição da lista de variáveis são configuradas na guia Geral na caixa de diálogo Opções (menu Editar, Opções).

Total simples para uma variável única

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.
3. Clique com o botão direito em *Categoria de idade* na área de janela de tela e escolha **estatísticas básicas** no menu pop-up.
4. Na caixa de diálogo estatísticas básicas, selecione **% de N da coluna** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
5. Na célula Rótulo na lista Exibição, exclua o rótulo padrão e digite Percentual.
6. Clique em **Aplicar à seleção**.
7. Clique com o botão direito em *Categoria de idade* na área de janela de tela e escolha **Categorias e totais** no menu pop-up.
8. Selecione (clique) **Total** na caixa de diálogo Categorias e totais.
9. Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	Total	2828	100.0%

Figura 20. Total simples para uma variável categórica única

O que você vê é o que será totalizado

Os totais são baseados em categorias exibidas na tabela. Se você optar por excluir algumas categorias de uma tabela, os casos dessas categorias não serão incluídos nos cálculos de totais.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).

2. Clique com o botão direito em *Categoria de idade* na área de janela de tela e escolha **Categorias e totais** no menu pop-up.
3. Clique na categoria rotulada *Menor que 25* na lista Rótulo.
4. Clique na tecla de seta à esquerda da lista Exclusão.
5. Clique na categoria rotulada *65 ou mais* na lista Rótulo.
6. Clique na tecla de seta à esquerda da lista Exclusão novamente.
As duas categorias são movidas da lista Exibição para a lista Exclusão.
7. Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Count	Percent
Age category	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%
	Total	2107	100.0%

Figura 21. Total na tabela com categorias excluídas

A contagem total nessa tabela é de apenas 2.107, comparada com 2.828 quando todas as categorias são incluídas. Apenas as categorias que são usadas na tabela são incluídas no total. (A porcentagem total ainda é 100%, porque todas as porcentagens são baseadas no número total de casos usados na tabela, não no número total de casos no arquivo de dados.)

Posição de exibição de totais

Por padrão, os totais são exibidos abaixo das categorias que estão sendo totalizadas. É possível mudar a posição de exibição de totais para mostrá-los acima das categorias que estão sendo totalizadas.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique com o botão direito em *Categoria de idade* na área de janela de tela e escolha **Categorias e totais** no menu pop-up.
3. No grupo Aparecimento de totais e subgrupos, selecione **Acima das categorias às quais eles se aplicam**.
4. Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Count	Percent
Age category	Total	2107	100.0%
	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%

Figura 22. Total exibido acima de categorias totalizadas

Totais para tabelas aninhadas

Como os totais podem ser aplicados a variáveis categóricas em qualquer nível do aninhamento, é possível criar tabelas que contêm os totais de grupo em vários níveis de aninhamento.

Totais de grupo

Os totais para variáveis categóricas aninhadas em outras variáveis categóricas representam totais de grupo.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Arraste e solte *Gênero* à esquerda de *Categoria de idade* na área de janela de tela.

- Clique com o botão direito em *Categoria de idade* na área de janela de tela e escolha **Categorias e totais** no menu pop-up.
Antes de criar a tabela, vamos mover os totais de volta abaixo das categorias totalizadas.
- No grupo Aparecimento de totais e subgrupos, selecione **Abaixo das categorias às quais elas se aplicam**.
- Clique em **Aplicar** para salvar a configuração e retornar ao construtor de tabela.
- Clique em **OK** para criar a tabela.

				Count	Percent
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			Total	942	100.0%
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
			Total	1165	100.0%

Figura 23. Totais de categoria de idade em categorias de Gênero

A tabela agora exibe dois totais de grupo: um para homens e um para mulheres.

Totais gerais

Os totais aplicados a variáveis aninhadas são sempre totais de grupo, não totais gerais. Se desejar totais para a tabela inteira, é possível aplicar totais à variável no nível de aninhamento mais externo.

- Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
- Clique com o botão direito em *Gênero* na área de janela de tela e escolha **Categoria e totais** no menu pop-up.
- Selecione (clique) **Total** na caixa de diálogo Categorias e totais.
- Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

				Count	Percent		
Gender	Male	Age category	25 to 34	276	29.3%		
			35 to 44	309	32.8%		
			45 to 54	221	23.5%		
			55 to 64	136	14.4%		
			Total	942	100.0%		
	Female	Age category	25 to 34	351	30.1%		
			35 to 44	370	31.8%		
			45 to 54	260	22.3%		
			55 to 64	184	15.8%		
			Total	1165	100.0%		
			Total	Age category	25 to 34	627	29.8%
					35 to 44	679	32.2%
	45 to 54	481			22.8%		
	55 to 64	320			15.2%		
	Total	2107			100.0%		

Figura 24. Totais gerais para uma tabela aninhada

Observe que o total geral é de apenas 2.107, não 2.828. Duas categorias de idade ainda são excluídas da tabela, portanto, os casos nestas categorias são excluídos de todos os totais.

Totais de variáveis de estrato

Os totais para variáveis de estrato são exibidos como estratos separadas na tabela.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em **Camadas** no construtor de tabela para exibir a lista Camadas.
3. Arraste e solte *Gênero* da área de linha na área de janela de tela para a lista Camadas.
Nota: Como você já especificou totais para *Gênero*, não é necessário fazer isso agora. Mover a variável entre dimensões não afeta nenhuma das configurações para essa variável.
4. Clique em **OK** para criar a tabela.
5. Clique duas vezes na tabela no Visualizador para ativá-lo.
6. Clique na seta para baixo na lista suspensa Camada para exibir uma lista de todas as estratos na tabela.

Existem três estratos na tabela: *Gênero masculino*, *Gênero feminino* e *Total de gêneros*.

Posição de exibição de totais de estratos

Para totais de variáveis de estrato, a posição de exibição (acima ou abaixo) para totais determina a posição da estrato para os totais. Por exemplo, se você especificar **Acima das categorias às quais eles se aplicam** para um total de variáveis de estrato, a estrato total será a primeira estrato exibida.

Subtotais

É possível incluir subtotais para subconjuntos de categorias de uma variável. Por exemplo, é possível incluir subtotais para categorias de idade que representam todos os respondentes na pesquisa de opinião de amostra com idades abaixo e acima de 45.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em **Reconfigurar** para limpar todas as configurações anteriores no construtor de tabela.
3. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.
4. Clique com o botão direito em *Categoria de idade* na área de janela de tela e escolha **Categorias e totais** no menu pop-up.
5. Selecione **3,00** na lista Valor(es).
6. Clique em **Incluir subtotal** para exibir a caixa de diálogo Definir subtotal.
7. No campo de texto Rótulo, digite Subtotal < 45.
8. Em seguida, clique em **Continuar**.
Isso insere uma linha contendo o subtotal para as primeiras três categorias de idade.
9. Selecione **6,00** na lista Valor(es).
10. Clique em **Incluir subtotal** para exibir a caixa de diálogo Definir subtotal.
11. No campo de texto Rótulo, digite Subtotal 45+.
12. Em seguida, clique em **Continuar**.

Nota importante: Você deve selecionar a posição de exibição para totais e subtotais (**Acima das categorias às quais eles se aplicam** ou **Abaixo das categorias às quais eles se aplicam**) antes de definir quaisquer subtotais. A mudança da posição de exibição afeta todos os subtotais (não apenas o subtotal selecionado atualmente), e também *muda as categorias incluídas nos subtotais*.

13. Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Subtotal < 45	1548
	45 to 54	481
	55 to 64	320
	65 or older	479
	Subtotal 45+	1280

Figura 25. Subtotais para categoria de Idade

O que você vê é o que será subtotalizado

Assim como os totais, os subtotais são baseados nas categorias incluídas na tabela.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique com o botão direito em *Categoria de idade* na área de janela de tela e escolha **Categorias e totais** no menu pop-up.

Nota: O valor (não o rótulo) exibido para o primeiro subtotal é **1,00...3,00**, indicando que o subtotal inclui todos os valores na lista entre 1 e 3.

3. Selecione **1,00** na lista Valor(es) (ou clique no rótulo *Menor que 25*).
4. Clique na tecla de seta à esquerda da lista Exclusão.

A primeira categoria de idade agora é excluída e o valor exibido para o primeiro subtotal muda para **2,00...3,00**, indicando o fato de que a categoria excluída não será incluída no subtotal porque os subtotais são baseados nas categorias incluídas na tabela. A exclusão de uma categoria a exclui automaticamente de todos os subtotais, portanto, não é possível, por exemplo, exibir apenas subtotais sem as categorias nas quais os subtotais são baseados.

Ocultando o subtotal das categorias

É possível suprimir a exibição das categorias que definem um subtotal e exibir somente o subtotal, efetivamente "reduzindo" categorias sem afetar os dados subjacentes.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em **Reconfigurar** para limpar todas as configurações anteriores no construtor de tabela.
3. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.
4. Clique com o botão direito em *Categoria de idade* na área de janela de tela e escolha **Categorias e totais** no menu pop-up.
5. Selecione **3.00** na lista Valor(es).
6. Clique em **Incluir subtotal** para exibir a caixa de diálogo Definir subtotal.
7. No campo de texto Rótulo, digite *Menor que 45*.
8. Selecione (marque) **Ocultar subtotal das categorias da tabela**.
9. Em seguida, clique em **Continuar**.
Isso insere uma linha contendo o subtotal para as primeiras três categorias de idade.
10. Selecione **6,00** na lista Valor(es).
11. Clique em **Incluir subtotal** para exibir a caixa de diálogo Definir subtotal.
12. No campo de texto Rótulo, digite *45 ou mais*.
13. Selecione (marque) **Ocultar subtotal das categorias**.
14. Em seguida, clique em **Continuar**.
15. Para incluir um total com os subtotais, selecione (marque) **Total** no grupo Mostrar.
16. Clicar em **Aplicar**.

A tela reflete o fato de que os subtotais serão exibidos, mas as categorias que definem os subtotais serão excluídas.

17. Clique em **OK** para produzir a tabela.

		Count
Age category	Less than 45	1548
	45 or older	1280
	Total	2828

Figura 26. Tabela que exhibe apenas subtotais e totais

Subtotais de variáveis de estrato

Assim como os totais, os subtotais para variáveis de estrato são exibidos como estratos separadas na tabela. Basicamente, os subtotais são tratados como categorias. Cada categoria é uma estrato separada na tabela, e a ordem de exibição das categorias da estrato é determinada pela ordem da categoria especificada na caixa de diálogo Categorias e totais, incluindo a posição de exibição das categorias de subtotal.

Categorias calculadas para variáveis categóricas

É possível incluir categorias calculadas em tabelas customizadas. Essas são novas categorias que são calculadas de categorias da mesma variável em qualquer nível de aninhamento em qualquer dimensão -- linha, coluna ou camada. Por exemplo, é possível incluir uma categoria calculada que mostra a diferença entre duas categorias.

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *survey_sample.sav*. Consulte o tópico arquivos de dados para obter mais informações

Categoria calculada simples

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.
3. Clique com o botão direito em *Categoria de idade* na área de janela de tela e escolha **Categorias e totais** no menu pop-up.
4. Selecione **3.00** na lista Valor(es).
5. Clique em **Incluir categoria** para exibir a caixa de diálogo Definir categoria de cálculo.
6. No campo de texto Rótulo para categoria calculada, digite **Menor que 45**.
7. Selecione **Menor que 25 (1,00)** na lista Categorias e clique no botão de seta para copiá-la para a caixa de texto Expressão para categoria calculada. [1] é exibido na expressão.
8. Clique no botão do operador de sinal de mais (+) na caixa de diálogo (ou pressione a tecla + no teclado).
9. Selecione **25 a 34 (2,00)** na lista Categorias e clique no botão de seta para copiá-la para a caixa de texto Expressão para categoria calculada.
10. Clique no botão do operador de sinal de mais (+) na caixa de diálogo (ou pressione a tecla + no teclado).
11. Selecione **35 a 44 (3,00)** na lista Categorias e clique no botão de seta para copiá-la para a caixa de texto Expressão para categoria calculada.
12. Em seguida, clique em **Continuar**.

Isso insere uma linha contendo o subtotal para as primeiras três categorias de idade.

13. Selecione **5,00** na lista Valor(es).
14. Clique em **Incluir subtotal** para exibir a caixa de diálogo Definir subtotal.
15. No campo de texto Rótulo, digite Menor que 65.
16. Em seguida, clique em **Continuar**.

Isso insere uma linha contendo o subtotal para as cinco primeiras categorias.

17. Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

Age category	Count
Less than 25	242
25 to 34	627
35 to 44	679
Less than 45	1548
45 to 54	481
55 to 64	320
Less than 65	2349
65 or older	479

Figura 27. Categoria calculada com subtotal

A tabela inclui uma categoria calculada (*Menor que 45*) e um subtotal (*Menor que 65*). O subtotal inclui categorias também incluídas na categoria calculada. Não foi possível criar a mesma tabela somente com subtotais, porque os subtotais não podem compartilhar as mesmas categorias.

Ocultando categorias em uma categoria calculada

Assim como com subtotais, é possível suprimir a exibição das categorias que são usadas em uma expressão de categoria calculada e exibir somente a própria categoria calculada. O exemplo a seguir é construído sobre o anterior.

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. Clique com o botão direito em *Categoria de idade* na área de janela de tela e escolha **Categorias e totais** no menu pop-up.
3. Selecione a categoria calculada *Menor que 45* na lista Valor(es).
4. Clique em **Editar** para exibir a caixa de diálogo Definir categoria de cálculo.
5. Selecione **Ocultar categorias usadas em expressão da tabela**.
6. Em seguida, clique em **Continuar**.
7. Selecione o subtotal *Menor que 65* na lista Valor(es).
8. Clique em **Editar** para exibir a caixa de diálogo Definir subtotal.
9. Selecione **Ocultar subtotal das categorias da tabela**.
10. Em seguida, clique em **Continuar**.
11. Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

Age category	Count
Less than 45	1548
Less than 65	2349
65 or older	479

Figura 28. Categoria calculada com categorias de subtotal e ocultas

Como no exemplo anterior, a tabela inclui uma categoria calculada e um subtotal. Mas, nesse caso, as categorias em cada um ficam ocultas para que somente esses totais sejam mostrados.

Referenciando subtotais em uma categoria calculada

É possível incluir subtotais em uma expressão de categoria calculada.

- Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
- Clique em **Reconfigurar** para limpar todas as configurações anteriores no construtor de tabela.
- No construtor de tabela, arraste e solte *Status da força de mão de obra* da lista de variáveis para a área Linhas da área de janela de tela.
- Arraste e solte *Estado civil* da lista de variáveis para a área Colunas.
- Clique com o botão direito em *Status da força de mão de obra* na área de janela de tela e escolha **Categorias e totais** no menu pop-up.
- Selecione **2** na lista Valor(es).
- Clique em **Incluir subtotal** para exibir a caixa de diálogo Definir subtotal.
- No campo de texto Rótulo, digite *Trabalhando*.
- Selecione **Ocultar subtotal das categorias da tabela**.
- Em seguida, clique em **Continuar**.
Isso insere uma linha contendo o subtotal para as duas primeiras categorias de status de trabalho.
- Selecione **8** na lista Valor(es).
- Clique em **Incluir subtotal** para exibir a caixa de diálogo Definir subtotal.
- No campo de texto Rótulo, digite *Não trabalhando*.
- Selecione **Ocultar subtotal das categorias**.
- Em seguida, clique em **Continuar**.
Isso insere uma linha contendo o subtotal para as outras categorias de status de trabalho.
- Selecione o subtotal *Não trabalhando* na lista Valor(es).
- Clique em **Incluir categoria** para exibir a caixa de diálogo Definir categoria de cálculo.
- No campo de texto Rótulo para categoria calculada, digite *Trabalhando / Não trabalhando*.
- Selecione **Trabalhando (Trabalhando #1)** na lista Totais e subtotais e clique no botão de seta para copiá-lo para a caixa de texto Expressão para categoria calculada.
- Clique no botão do operador de divisão (/) na caixa de diálogo (ou pressione a tecla / no teclado).
- Selecione **Não trabalhando (Não trabalhando #2)** na lista Totais e subtotais e clique no botão de seta para copiá-lo para a caixa de texto Expressão para categoria calculada.
Por padrão, a categoria calculada usa o mesmo formato que a estatística da variável, que é Contagem nesse caso. Como queremos mostrar casas decimais resultantes da divisão na expressão da categoria calculada e o formato padrão para Contagem não inclui casas decimais, é necessário mudar o formato.
- Clique na guia Formatos de exibição.
- Mude a configuração Decimais para Contagem para **2**.
- Em seguida, clique em **Continuar**.
- Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Count	Count	Count	Count	Count
Labor force status	Working	916	64	330	67	494
	Not Working	429	219	116	26	169
	Working / Not Working	2.14	.29	2.84	2.58	2.92

Figura 29. Categoria calculada que mostra a razão de subtotais

A tabela inclui dois subtotais e uma categoria calculada. A categoria calculada mostra a razão dos subtotais para que seja possível comparar facilmente os grupos representados por cada subtotal. Há uma razão muito mais baixa de respondentes viúvos trabalhando para não trabalhando em comparação com os outros grupos. Além disso, há uma razão um pouco mais baixa de respondentes casados, talvez resultante de cônjuges que deixaram o emprego para ficar em casa com um filho.

Usando categorias calculadas para exibir subtotais não exaustivos

Os subtotais são exaustivos. Ou seja, todos os subtotais em uma tabela incluem todos os valores acima ou abaixo de sua posição na tabela. As categorias calculadas, por outro lado, não são exaustivas e permitem somar uma combinação de categorias em uma tabela.

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. Clique em **Reconfigurar** para limpar todas as configurações anteriores no construtor de tabela.
3. No construtor de tabela, arraste e solte *Pensar em si mesmo como liberal ou conservador* da lista de variáveis para a área linha da área de janela de tela.
4. Clique com o botão direito em *Pensar em si mesmo como liberal ou conservador* na área de janela de tela e escolha **Categorias e totais** no menu de contexto pop-up.
5. Selecione **3** na lista Valor(es).
6. Clique em **Incluir categoria** para exibir a caixa de diálogo Definir categoria calculada.
7. No campo de texto Rótulo para categoria calculada, digite Subtotal liberal. Observe que existem quatro espaços antes do texto. Estes espaços são usados para indentação na tabela resultante.
8. Selecione **Extremamente liberal (1)** na lista Categorias e clique no botão de seta para copiá-la para a caixa de texto Expressão para categoria calculada.
9. Clique no botão do operador de sinal de mais (+) na caixa de diálogo (ou pressione a tecla + no teclado).
10. Selecione **Liberal (2)** na lista Categorias e clique no botão de seta para copiá-la para a caixa de texto Expressão para categoria calculada.
11. Clique no botão do operador de sinal de mais (+) na caixa de diálogo (ou pressione a tecla + no teclado).
12. Selecione **Pouco liberal (3)** na lista Categorias e clique no botão de seta para copiá-la para a caixa de texto Expressão para categoria calculada.
13. Clique em **Continuar**.
Isso insere uma linha contendo o subtotal para categorias liberais.
14. Selecione **7** na lista Valor(es).
15. Clique em **Incluir categoria** para exibir a caixa de diálogo Definir categoria calculada.
16. No campo de texto Rótulo para categoria calculada, digite Subtotal conservador. Observe que existem quatro espaços antes do texto. Estes espaços são usados para indentação na tabela resultante.
17. Selecione **Pouco conservador (5)** na lista Categorias e clique no botão de seta para copiá-la para a caixa de texto Expressão para categoria calculada.
18. Clique no botão do operador de sinal de mais (+) na caixa de diálogo (ou pressione a tecla + no teclado).
19. Selecione **Conservador (6)** na lista Categorias e clique no botão de seta para copiá-la para a caixa de texto Expressão para categoria calculada.
20. Clique no botão do operador de sinal de mais (+) na caixa de diálogo (ou pressione a tecla + no teclado).
21. Selecione **Extremamente conservador (7)** na lista Categorias e clique no botão de seta para copiá-la para a caixa de texto Expressão para categoria calculada.
22. Clique em **Continuar**.
Isso insere uma linha contendo o subtotal para categorias conservadoras.

23. Clique em **Aplicar** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Count
Think of self as liberal or conservative	Extremely liberal	64
	Liberal	357
	Slightly liberal	351
	Liberal Subtotal	772
	Moderate	986
	Slightly conservative	432
	Conservative	415
	Extremely conservative	86
	Conservative Subtotal	933

Figura 30. Categorias calculadas que exibem subtotais não exaustivos

A tabela inclui duas categorias calculadas que não incluem todas as categorias exibidas na tabela. A categoria *Moderado* não está incluída em nenhuma categoria calculada. Não é possível criar a mesma tabela com subtotais porque os subtotais são exaustivos.

Tabelas para variáveis com categorias compartilhadas

As pesquisas de opinião geralmente contêm muitas perguntas com um conjunto comum de respostas possíveis. Por exemplo, nossa pesquisa de opinião de amostra contém diversas variáveis referentes a confiança em várias instituições e serviços públicos e privados, todas com o mesmo conjunto de categorias de resposta: 1 = *Uma grande quantidade*, 2 = *Somente alguns* e 3 = *Quase nenhum*. É possível usar o empilhamento para exibir essas variáveis relacionadas na mesma tabela -- e é possível exibir as categorias de resposta compartilhada nas colunas da tabela. Esses recursos também estarão disponíveis se você usar categorias calculadas, com a condição de que o rótulo e a expressão de qualquer categoria calculada sejam iguais em todas as variáveis.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	490	1068	306
Confidence in education	511	1055	315
Confidence in major companies	500	1078	243
Confidence in medicine	844	864	167
Confidence in press	176	878	808
Confidence in television	196	936	744

Figura 31. Tabela de variáveis com categorias compartilhadas

Nota: Na versão anterior de Tabelas customizadas, isso era conhecido como uma "tabela de frequências".

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *survey_sample.sav*. Consulte o tópico arquivos de dados para obter mais informações.

Todos os exemplos fornecidos aqui exibem rótulos de variáveis em caixas de diálogo, ordenados em ordem alfabética. As propriedades de exibição da lista de variáveis são configuradas na guia Geral na caixa de diálogo Opções (menu Editar, Opções).

Tabela de contagens

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. Na lista de variáveis no construtor de tabela, clique em *Confiança em bancos...* e, em seguida, Shift-clique em *Confiança na televisão* para selecionar todas as variáveis de "confiança". (Nota: Isso supõe que os rótulos de variáveis sejam exibidos em ordem alfabética, não em ordem de arquivo, na lista de variáveis.)
3. Arraste e solte as seis variáveis de confiança para a área Linhas na área de janela de tela. Isso empilha as variáveis na dimensão da linha. Por padrão, os rótulos de categoria para cada variável também são exibidos nas linhas, resultando em uma tabela muito longa e estreita (6 variáveis x 3 categorias = 18 linhas) -- mas, como todas as seis variáveis compartilham os mesmos rótulos de categoria definidos (rótulos de valor), é possível colocar os rótulos de categoria na dimensão da coluna.
4. Na lista suspensa Posição da categoria, selecione **Rótulos de linha em colunas**. Agora a tabela possui apenas seis linhas, uma para cada uma das variáveis empilhadas, e as categorias definidas tornam-se colunas na tabela.
5. Antes de criar a tabela, selecione (clique) **Ocultar** para Posição no grupo estatísticas básicas, pois o rótulo de estatística de sumarização *Contagem* não é realmente necessário.
6. Clique em **OK** para criar a tabela.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	490	1068	306
Confidence in education	511	1055	315
Confidence in major companies	500	1078	243
Confidence in medicine	844	864	167
Confidence in press	176	878	808
Confidence in television	196	936	744

Figura 32. Tabela de variáveis de linha empilhadas com rótulos de categoria compartilhados em colunas

Em vez de exibir as variáveis nas linhas e categorias nas colunas, é possível criar uma tabela com as variáveis empilhadas nas colunas e as categorias exibidas nas linhas. Essa pode ser uma opção melhor se houver mais categorias do que variáveis, embora em nosso exemplo haja mais variáveis do que categorias.

Tabela de porcentagens

Para uma tabela com variáveis empilhadas em linhas e categorias exibidas em colunas, a porcentagem mais significativa (ou pelo menos a mais fácil de entender) a ser exibida são porcentagens de linhas. (Para uma tabela com variáveis empilhadas nas colunas e categorias exibidas nas linhas, provavelmente você deseja porcentagens da coluna.)

1. Abra o construtor de tabela novamente (menu **Analisar, Tabelas, Tabelas customizadas**).
2. Clique com o botão direito em qualquer uma das variáveis de confiança na visualização da tabela na área de janela de tela e escolha **estatísticas básicas** no menu pop-up.
3. Selecione **% de N da linha** na lista Estatísticas e clique no botão de seta para movê-la para a lista Exibição.
4. Clique em qualquer célula na linha *Contagem*, na lista Exibição, e clique no botão de seta para movê-la de volta para a lista Estatísticas, removendo-a da lista Exibição.
5. Clique em **Aplicar a todos** para aplicar a mudança na estatística de sumarização a todas as variáveis empilhadas na tabela.

Nota: Se sua visualização da tabela não for parecida com essa figura, provavelmente você clicou em **Aplicar à seleção** em vez de **Aplicar a todos**, que aplica a nova estatística de sumarização apenas à variável selecionada. Nesse exemplo, isso resultaria em duas colunas para cada categoria: uma com itens temporários de contagem exibidos para todas as outras variáveis e uma com um item temporário de porcentagem de linhas exibido para a variável selecionada. Essa é exatamente a tabela que seria produzida, mas *não* a que desejamos nesse exemplo.

6. Clique em **OK** para criar a tabela.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	26.3%	57.3%	16.4%
Confidence in education	27.2%	56.1%	16.7%
Confidence in major companies	27.5%	59.2%	13.3%
Confidence in medicine	45.0%	46.1%	8.9%
Confidence in press	9.5%	47.2%	43.4%
Confidence in television	10.4%	49.9%	39.7%

Figura 33. Tabela de porcentagens de linhas para variáveis empilhadas em linhas, categorias exibidas em colunas

Nota: É possível incluir qualquer número de estatísticas básicas em uma tabela de variáveis com categorias compartilhadas. Nossos exemplos mostram somente uma por vez para mantê-las simples.

Controle de totais e de categorias

É possível criar tabelas com categorias na dimensão oposta das variáveis somente se todas as variáveis na tabela tiverem as mesmas categorias, exibidas na mesma ordem. Isso inclui totais, subtotais e quaisquer outros ajustamentos da categorias feitos. Isso significa que quaisquer modificações feitas na caixa de diálogo Categorias e totais devem ser feitas para todas as variáveis na tabela que compartilham as categorias.

1. Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
2. Clique com o botão direito na primeira variável de confiança na visualização da tabela na área de janela de tela e escolha **Categorias e totais** no menu pop-up.
3. Selecione (marque) **Total** na caixa de diálogo Categorias e totais e, em seguida, clique em **Aplicar**.
A primeira coisa que provavelmente você notará é que os rótulos de categoria foram movidos das colunas de volta para as linhas. Você também pode notar que o controle Posição da categoria agora está desativado. Isso ocorre porque as variáveis não compartilham mais exatamente o mesmo conjunto de "categorias". Uma das variáveis agora tem uma categoria total.
4. Clique com o botão direito em qualquer uma das variáveis de confiança na área de janela de tela e selecione **Selecionar todas as variáveis de linha** no menu pop-up -- ou Ctrl-clique em cada variável empilhada na área de janela de tela até que todas estejam selecionadas (pode ser necessário rolar para baixo na área de janela ou expandir a janela do construtor de tabela).
5. Clique em **Categorias e totais** no grupo Definição.
6. Se **Total** ainda não estiver selecionado (marcado) na caixa de diálogo Categorias e totais, selecione-o agora e, em seguida, clique em **Aplicar**.
7. A lista suspensa Posição da categoria deve ser ativada novamente, pois agora todas as variáveis possuem a categoria total adicional, portanto, selecione **Rótulos da linha em colunas**.
8. Clique em **OK** para criar a tabela.

	A great deal	Only some	Hardly any	Total
Confidence in banks & financial institutions	26.3%	57.3%	16.4%	100.0%
Confidence in education	27.2%	56.1%	16.7%	100.0%
Confidence in major companies	27.5%	59.2%	13.3%	100.0%
Confidence in medicine	45.0%	46.1%	8.9%	100.0%
Confidence in press	9.5%	47.2%	43.4%	100.0%
Confidence in television	10.4%	49.9%	39.7%	100.0%

Figura 34. Tabela de porcentagens de linhas para variáveis empilhadas em linhas, categorias e totais exibidos em colunas

Aninhando em tabelas com categorias compartilhadas

Em tabelas aninhadas, as variáveis empilhadas com as categorias compartilhadas devem estar no nível de aninhamento mais interno de sua dimensão, se você desejar exibir os rótulos de categoria na dimensão oposta.

1. Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
2. Arraste e solte *Gênero* na lista de variáveis para o lado esquerdo da área Linhas.
As variáveis empilhadas com categorias compartilhadas agora estão aninhadas em categorias de gênero na visualização da tabela.
3. Agora arraste e solte *Gênero* à direita de uma das variáveis de confiança empilhadas na visualização da tabela.

Mais uma vez, os rótulos de categoria foram revertidos para a dimensão da linha, e o controle Posição da categoria foi desativado. Agora você tem uma variável empilhada que também tem *Gênero* aninhado nela, enquanto as outras variáveis empilhadas não contêm variáveis aninhadas. É possível incluir *Gênero* como uma variável aninhada em cada uma das variáveis empilhadas, mas mover rótulos da linha para colunas resulta nos rótulos de categoria para *Gênero* sendo exibidos nas colunas, não os rótulos de categoria para as variáveis empilhadas com as categorias compartilhadas. Isso ocorre porque *Gênero* agora seria a variável aninhada mais interna, e a mudança da posição de categoria sempre se aplica à variável aninhada mais interna.

Estatísticas básicas

As **estatísticas básicas** incluem tudo, de contagens simples para variáveis categóricas a medidas de dispersão, como o erro padrão da média para variáveis de escala. Elas *não* incluem testes de significância disponíveis na guia Estatísticas do teste na caixa de diálogo Tabelas customizadas. Consulte o tópico “Estatísticas de teste” na página 55 para obter mais informações

As estatísticas básicas para variáveis categóricas e conjuntos de múltiplas respostas incluem contagens e uma grande variedade de cálculos de porcentagem, incluindo:

- Porcentagens de linhas
- Porcentagens da coluna
- Porcentagens de subtabela
- Porcentagens da tabela
- Porcentagens de N válido

Além das estatísticas básicas disponíveis para variáveis categóricas, as estatísticas básicas para variáveis de escala e sumarizações de total customizado para variáveis categóricas incluem:

- Média

- Mediana
- Percentis
- Soma
- Desvio padrão
- Intervalo
- Valores mínimo e máximo

As estatísticas básicas adicionais estão disponíveis para conjuntos de múltiplas respostas. Uma lista completa de estatísticas básicas também está disponível. Consulte o tópico “Estatísticas básicas” na página 6 para obter mais informações

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *survey_sample.sav*. Consulte o tópico arquivos de dados para obter mais informações.

Todos os exemplos fornecidos aqui exibem rótulos de variáveis em caixas de diálogo, ordenados em ordem alfabética. As propriedades de exibição da lista de variáveis são configuradas na guia Geral na caixa de diálogo Opções (menu Editar, Opções).

Variável de origem estatística de sumarização

As estatísticas básicas disponíveis dependem do nível de medição da variável de origem estatística de sumarização. A origem das estatísticas básicas (a variável na qual as estatísticas básicas são baseadas) é determinada por:

- **Nível de medição.** Se uma tabela (ou uma seção da tabela em uma tabela empilhada) contiver uma variável de escala, as estatísticas básicas serão baseadas na variável de escala.
- **Ordem de seleção de variáveis.** A dimensão de origem estatística padrão (linha ou coluna) para variáveis categóricas é baseada na ordem em que você arrasta e solta variáveis na área de janela de tela. Por exemplo, se você arrastar uma variável para a área de linhas primeiro, a dimensão da linha será a dimensão de origem estatística padrão.
- **Aninhamento.** Para variáveis categóricas, as estatísticas básicas são baseadas na variável mais interna na dimensão de origem estatística.

Uma tabela empilhada pode ter diversas variáveis de origem estatística de sumarização (de escala e categóricas), mas cada seção da tabela possui apenas uma origem estatística de sumarização.

Origem de estatísticas básicas para variáveis categóricas

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas da área de janela de tela.
3. Clique com o botão direito em *Categoria de idade* na área de janela de tela e selecione **estatísticas básicas** no menu pop-up. (Como essa é a única variável na tabela, ela é a variável de origem estatística.)
4. Na caixa de diálogo estatísticas básicas, selecione *% de N da coluna* na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
5. Clique em **Aplicar à seleção**.
6. No construtor de tabela, arraste e solte *Obter notícias da Internet*, à direita de *Categoria de idade* na área de janela de tela.
7. Clique com o botão direito em *Categoria de idade* na área de janela de tela novamente. O item **estatísticas básicas** no menu pop-up agora está desativado, porque a *Categoria de idade* não é a variável aninhada mais interna na dimensão de origem estatística.

8. Clique com o botão direito em *Obter notícias da Internet* na área de janela de tela. O item **estatísticas básicas** está ativado, porque agora ele é a variável de origem estatística de sumarização, já que é a variável aninhada mais interna na dimensão de origem estatística. (Como a tabela possui linhas apenas de uma dimensão — ela é a dimensão de origem estatística.)
9. Arraste e solte *Obter notícias da Internet* da área Linhas na área de janela de tela para a área Colunas.
10. Clique com o botão direito *Obter notícias da Internet* na área de janela de tela novamente. O item **estatísticas básicas** no menu pop-up agora está desativado, porque a variável não está mais na dimensão de origem estatística.

Categoria de idade é mais uma vez a variável de origem estatística, porque a dimensão de origem estatística padrão para variáveis categóricas é a primeira dimensão na qual você coloca variáveis ao criar a tabela. Nesse exemplo, a primeira coisa que fizemos foi colocar variáveis na dimensão da linha. Portanto, a dimensão da linha é a dimensão de origem estatística padrão; e como a *Categoria de idade* agora é a única variável nessa dimensão, ela é a variável de origem estatística.

Origem de estatísticas básicas para variáveis de escala

1. Arraste e solte a variável de escala *Horas por dia assistindo à TV* à esquerda de *Categoria de idade* na área Linhas da área de janela de tela.
A primeira coisa que você pode notar é que as sumarizações *Contagem* e *% de N da coluna* foram substituídas por *Média* -- e se você clicar com o botão direito em *Horas por dia assistindo à TV* na área de janela de tela, verá que agora essa opção agora é a variável de origem estatística de sumarização. Para uma tabela com uma variável de escala, a variável de escala é sempre a variável de origem estatística, independentemente de seu nível de aninhamento ou dimensão, e a estatística de sumarização padrão para variáveis de escala é a média.
2. Arraste e solte *Horas por dia assistindo à TV* da área Linhas para a área Colunas, acima de *Obter notícias da Internet*.
3. Clique com o botão direito em *Horas por dia assistindo à TV* e selecione **estatísticas básicas** no menu pop-up. (Ela ainda é a variável de origem estatística, mesmo quando é movida para uma dimensão diferente.)
4. Na caixa de diálogo estatísticas básicas, clique na célula **Formato** para a média na lista Exibição e selecione **nnnn** na lista suspensa Formato. (Pode ser necessário rolar para cima na lista para localizar essa opção.)
5. Na célula Casas Decimais, digite 2.
6. Clique em **Aplicar à seleção**.
A visualização da tabela na área de janela de tela agora mostra que os valores médios serão exibidos com dois decimais.
7. Clique em **OK** para criar a tabela.

		Hours per day watching TV	
		Get news from internet	
		No	Yes
		Mean	Mean
Age category	Less than 25	3.54	2.12
	25 to 34	3.42	2.14
	35 to 44	3.00	2.01
	45 to 54	2.83	2.06
	55 to 64	3.24	2.37
	65 or older	3.82	2.33

Figura 35. Variável de escala sumarizada em variáveis categóricas de tabulação cruzada

Variáveis empilhadas

Como uma tabela empilhada pode conter várias variáveis de origem estatística e é possível especificar diferentes estatísticas básicas para cada uma dessas variáveis de origem estatística, existem algumas considerações especiais para especificar estatísticas básicas em tabelas empilhadas.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em **Reconfigurar** para limpar todas as configurações anteriores no construtor de tabela.
3. Clique em *Obter notícias da Internet* na lista de variáveis e, em seguida, shift-clique em *Obter notícias da televisão* na lista de variáveis para selecionar todas as variáveis de "notícias". (*Nota: Isso supõe que os rótulos de variáveis sejam exibidos em ordem alfabética, não em ordem de arquivo, na lista de variáveis.*)
4. Arraste e solte as cinco variáveis de notícias na área Linhas da área de janela de tela.
As cinco variáveis de notícias são empilhadas na dimensão da linha.
5. Clique em *Obter notícias da Internet* na área de janela de tela para que apenas essa variável seja selecionada.
6. Agora clique com o botão direito em *Obter notícias da Internet* e selecione **estatísticas básicas** no menu pop-up.
7. Na caixa de diálogo estatísticas básicas, selecione *% de N da coluna* na lista Estatísticas e clique na seta para incluí-la na lista Exibição. (É possível usar a seta para mover as estatísticas selecionadas da lista Estatísticas para a lista Exibição, ou arrastar e soltar estatísticas selecionadas da lista Estatísticas para a lista Exibição.)
8. Em seguida, clique em **Aplicar à seleção**.
Uma coluna é incluída para porcentagens da coluna -- mas a visualização da tabela na área de janela de tela indica que as porcentagens da coluna serão exibidas apenas para uma variável. Isso ocorre porque, em uma tabela empilhada, existem diversas variáveis de origem estatística, e cada uma pode ter diferentes estatísticas básicas. No entanto, nesse exemplo, desejamos exibir as mesmas estatísticas básicas para todas as variáveis.
9. Clique com o botão direito em *Obter notícias de jornais*, na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
10. Na caixa de diálogo estatísticas básicas, selecione *% de N da coluna* na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
11. Em seguida, clique em **Aplicar a todos**.

Agora a visualização da tabela indica que as porcentagens da coluna serão exibidas para todas as variáveis empilhadas.

Total de estatísticas básicas customizada para variáveis categóricas

Para variáveis de origem estatística categóricas, é possível incluir estatísticas básicas total customizada que são diferentes das estatísticas exibidas para as categorias da variável. Por exemplo, para uma variável ordinal, é possível exibir porcentagens para cada categoria e a média ou mediana para a estatística de sumarização total customizada.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em **Reconfigurar** para limpar todas as configurações anteriores no construtor de tabela.
3. Clique em *Confiança na imprensa* na lista de variáveis e, em seguida, Ctrl-clique em *Confiança na TV* para selecionar ambas as variáveis.
4. Arraste e solte as duas variáveis na área Linhas da área de janela de tela. Isso empilha as duas variáveis na dimensão da linha.
5. Clique com o botão direito em qualquer variável na área de janela de tela e selecione **Selecionar todas as variáveis de linha** no menu pop-up. (Ambas podem já estar selecionadas, mas queremos ter certeza.)
6. Clique com o botão direito na variável novamente e selecione **Categorias e totais** no menu pop-up.

7. Na caixa de diálogo Categorias e totais, clique em (marque) **Total** e, em seguida, clique em **Aplicar**.
A visualização da tabela na área de janela de tela agora exibe uma linha total para ambas as variáveis. Para exibir estatísticas básicas total customizada, os totais e/ou subtotais devem ser especificados para a tabela.
8. Clique com o botão direito em qualquer variável na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
9. Na caixa de diálogo estatísticas básicas, clique em *Contagem* na lista Exibição e clique na seta para movê-la para a lista Estatísticas, removendo-a da lista Exibição.
10. Clique em *% de N da coluna* na lista Estatísticas e clique na tecla de seta para movê-la para a lista Exibição.
11. Clique em (marque) **estatísticas básicas customizada para totais e subtotais**.
12. Clique em *Contagem* na lista Exibição de sumarização customizada e clique na seta para movê-la para a lista estatísticas básicas customizada, removendo-a da lista Exibição.
13. Clique em *Média* na lista estatísticas básicas customizada e clique na seta para movê-la para a lista Exibição de sumarização customizada.
14. Clique na célula **Formato** para a média na lista Exibição e selecione **nnnn** na lista suspensa de formatos. (Pode ser necessário rolar para cima na lista para localizar essa opção.)
15. Na célula Casas decimais, digite 2.
16. Clique em **Aplicar a todos** para aplicar essas configurações a ambas as variáveis na tabela.
Uma nova coluna foi incluída para a estatística de sumarização total customizada, que pode não ser o que você deseja, pois a visualização na área de janela de tela indica claramente que isso resultará em uma tabela com muitas células vazias.
17. No construtor de tabela, no grupo estatísticas básicas, selecione **Linhas** na lista suspensa Posição.
Isso move todas as estatísticas básicas para a dimensão da linha, exibindo todas as estatísticas básicas em uma única coluna na tabela.
18. Clique em **OK** para criar a tabela.

Confidence in press	A great deal	Column N %	9.5%
	Only some	Column N %	47.2%
	Hardly any	Column N %	43.4%
	Total	Mean	2.34
Confidence in television	A great deal	Column N %	10.4%
	Only some	Column N %	49.9%
	Hardly any	Column N %	39.7%
	Total	Mean	2.29

Figura 36. Variáveis categóricas com as estatísticas básicas total customizada

Exibindo valores de categoria

Há somente um pequeno problema com a tabela anterior -- pode ser difícil interpretar o valor médio sem saber os valores da categoria subjacente nos quais ele é baseado. Uma média de 2,34 está em algum lugar entre *Uma grande quantidade* e *Somente alguns* -- ou está em algum lugar entre *Somente alguns* e *Quase nenhum*?

Embora não seja possível tratar esse problema diretamente nas Tabelas customizadas, é possível tratá-lo de uma forma mais geral.

1. Nos menus, escolha:
Editar > Opções...
2. Na caixa de diálogo Opções, clique na guia **Rótulos de saída**.
3. No grupo Rotulagem de tabela dinâmica, selecione **Valores e rótulos** da lista suspensa **Valores de variáveis em rótulos mostrados como**.

- Clique em **OK** para salvar esta configuração.
- Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas) e clique em **OK** para criar a tabela novamente.

Confidence in press	1 A great deal	Column N %	9.5%
	2 Only some	Column N %	47.2%
	3 Hardly any	Column N %	43.4%
	Total	Mean	2.34
Confidence in television	1 A great deal	Column N %	10.4%
	2 Only some	Column N %	49.9%
	3 Hardly any	Column N %	39.7%
	Total	Mean	2.29

Figura 37. Valores e rótulos exibidos para categorias da variável

Os valores de categoria deixam claro que uma média de 2,34 está em algum lugar entre *Somente alguns* e *Quase nenhum*. A exibição de valores de categoria na tabela facilita muito mais a interpretação do valor de estatísticas básicas total customizada, como a média.

Esta configuração de exibição é uma configuração global que afeta todas as saídas de tabela dinâmica de todos os procedimentos e persiste entre as sessões até que seja mudada. Para mudar a configuração de volta para exibir somente rótulos de valor:

- Nos menus, escolha:
Editar > Opções...
- Na caixa de diálogo Opções, clique na guia **Rótulos de saída**.
- No grupo Rótulo de tabela dinâmica, selecione **Rótulos** da lista suspensa **Valores da variável em rótulos mostrados como**.
- Clique em **OK** para salvar esta configuração.

Sumarizando variáveis de escala

Sumarizando variáveis de escala

Um amplo intervalo de estatísticas básicas está disponível para variáveis de escala. Além das contagens e porcentagens disponíveis para variáveis categóricas, as estatísticas básicas para variáveis de escala também incluem:

- Média
- Mediana
- Percentis
- Soma
- Desvio padrão
- Intervalo
- Valores mínimo e máximo

Consulte o tópico “estatísticas básicas para variáveis de escala e totais customizados categóricos” na página 9 para obter mais informações

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *survey_sample.sav*. Consulte o tópico arquivos de dados para obter mais informações

Todos os exemplos fornecidos aqui exibem rótulos de variáveis em caixas de diálogo, ordenados em ordem alfabética. As propriedades de exibição da lista de variáveis são especificadas na guia Geral na caixa de diálogo Opções (menu Editar, Opções).

Variáveis de escala empilhadas

É possível sumarizar diversas variáveis de escala na mesma tabela, empilhando-as na tabela.

1. Nos menus, escolha:

Analisar > Tabelas > Tabelas Customizadas...

2. No construtor de tabela, clique em *Idade do respondente* na lista de variáveis, Ctrl-clique em *Maior ano escolar concluído* e Ctrl-clique em *Horas por dia assistindo à TV* para selecionar todas as três variáveis.

3. Arraste e solte as três variáveis selecionadas na área Linhas da área de janela de tela.

As três variáveis são empilhadas na dimensão da linha. Como todas as três variáveis são variáveis de escala, nenhuma categoria é exibida e a estatística de sumarização padrão é a média.

4. Clique em **OK** para criar a tabela.

	Mean
Age of respondent	46
Highest year of school completed	13
Hours per day watching TV	3

Figura 38. Tabela de valores médios de variáveis de escala empilhadas

Várias estatísticas básicas

Por padrão, a média é exibida para variáveis de escala; no entanto, é possível escolher outras estatísticas básicas para variáveis de escala e exibir mais de uma estatística de sumarização.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).

2. Clique com o botão direito em qualquer uma das três variáveis de escala na visualização da tabela na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.

3. Na caixa de diálogo estatísticas básicas, selecione *Mediana* na lista Estatísticas e clique na seta para incluí-la na lista Exibição. (É possível usar a seta para mover as estatísticas selecionadas da lista Estatísticas para a lista Exibição, ou arrastar e soltar estatísticas selecionadas da lista Estatísticas para a lista Exibição.)

4. Clique na célula **Formatar** para a mediana na lista Exibição e selecione **nnnn** na lista suspensa de formatos.

5. Na célula Casas Decimais, digite 1.

6. Faça as mesmas mudanças para a média na lista Exibição.

7. Clique em **Aplicar a todos** para aplicar essas mudanças a todas as três variáveis de escala.

8. Clique em **OK** no construtor de tabela para criar a tabela.

	Mean	Median
Age of respondent	45.6	42.0
Highest year of school completed	13.3	13.0
Hours per day watching TV	2.9	2.0

Figura 39. Média e mediana exibidas na tabela de variáveis de escala empilhadas

Contagem, N válido e Valores omissos

Geralmente é útil exibir o número de casos usados para calcular estatísticas básicas, como a média, e é possível considerar (não de forma irracional) que a *Contagem* das estatísticas básicas fornece essas informações. No entanto, isso não fornecerá uma base de caso precisa se houver valores omissos. Para obter uma base de caso precisa, use *N válido*.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique com o botão direito em qualquer uma das três variáveis de escala na visualização da tabela na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
3. Na caixa de diálogo estatísticas básicas, selecione **Contagem** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
4. Em seguida, selecione **N válido** na lista Estatísticas e clique na seta para incluí-lo na lista Exibição.
5. Clique em **Aplicar a todos** para aplicar essas mudanças a todas as três variáveis de escala.
6. Clique em **OK** no construtor de tabela para criar a tabela.

	Mean	Median	Count	Valid N
Age of respondent	45.6	42.0	2832	2828
Highest year of school completed	13.3	13.0	2832	2820
Hours per day watching TV	2.9	2.0	2832	2337

Figura 40. Contagem versus N válido

Por todas as três variáveis, a *Contagem* é a mesma: 2.832. Não por coincidência, este é o número total de casos no arquivo de dados. Como as variáveis de escala não estão aninhadas em nenhuma variável categórica, a *Contagem* simplesmente representa o número total de casos no arquivo de dados.

Por outro lado, *N válido* é diferente para cada variável e difere muito da *Contagem* para *Horas por dia assistindo à TV*. Isso ocorre porque há uma grande quantidade de **valores omissos** para essa variável -- ou seja, os casos com nenhum valor registrado para essa variável ou valores definidos como representando dados omissos (como um código de 99 para representar *Não aplicável* para gravidez em homens).

7. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
8. Clique com o botão direito em qualquer uma das três variáveis de escala na visualização da tabela na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
9. Na caixa de diálogo estatísticas básicas, selecione **N válido** na lista Exibição e clique na tecla de seta para movê-lo de volta para a lista Estatísticas, removendo-o da lista Exibição.
10. Selecione **Contagem** na lista Exibição e clique na tecla de seta para movê-la de volta para a lista Estatísticas, removendo-a da lista Exibição.
11. Selecione **Omisso** na lista Estatísticas e clique na tecla de seta para incluí-lo na lista Exibição.
12. Clique em **Aplicar a todos** para aplicar essas mudanças a todas as três variáveis de escala.
13. Clique em **OK** no construtor de tabela para criar a tabela.

	Mean	Median	Missing
Age of respondent	45.6	42.0	4
Highest year of school completed	13.3	13.0	12
Hours per day watching TV	2.9	2.0	495

Figura 41. Número de valores omissos exibidos na tabela de estatísticas básicas de escala

A tabela agora exibe o número de valores omissos para cada variável de escala. Isso torna bem evidente que *Horas por dia assistindo à TV* tem um grande número de valores omissos, enquanto as outras duas

variáveis têm pouquíssimos valores. Esse pode ser um fator a ser considerado antes de colocar muita confiança nos valores de sumarização para essa variável.

Sumarizações diferentes para variáveis diferentes

Além de exibir várias estatísticas básicas, é possível exibir diferentes estatísticas básicas para variáveis de escala diferentes em uma tabela empilhada. Por exemplo, a tabela anterior revelou que apenas uma das três variáveis tem um grande número de valores omissos; portanto, talvez você queira mostrar o número de valores omissos apenas para essa variável.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em *Idade do respondente* na visualização da tabela na área de janela de tela e, em seguida, Ctrl-clique em *Maior ano escolar concluído* para selecionar ambas as variáveis.
3. Clique com o botão direito em qualquer uma das duas variáveis selecionadas e selecione **estatísticas básicas** no menu pop-up.
4. Na caixa de diálogo estatísticas básicas, selecione **Omisso** na lista Exibição e clique na tecla de seta para movê-lo de volta para a lista Estatísticas, removendo-o da lista Exibição.
5. Clique em **Aplicar à seleção** para aplicar a mudança apenas às duas variáveis selecionadas.

Os itens temporários nas células de dados da tabela indicam que o número de valores omissos será exibido apenas para *Horas por dia assistindo à TV*.

6. Clique em **OK** para criar a tabela.

	Mean	Median	Missing
Age of respondent	45.6	42.0	
Highest year of school completed	13.3	13.0	
Hours per day watching TV	2.9	2.0	495

Figura 42. Tabela de diferentes estatísticas básicas para variáveis diferentes

Embora essa tabela forneça as informações desejadas, o layout pode dificultar a interpretação da tabela. Alguém que está lendo a tabela pode achar que as células em branco na coluna *Omisso* indicam zero valores omissos para essas variáveis.

7. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
8. No grupo estatísticas básicas no construtor de tabela, selecione **Linhas** na lista suspensa Posição.
9. Clique em **OK** para criar a tabela.

Age of respondent	Mean	45.6
	Median	42.0
Highest year of school completed	Mean	13.3
	Median	13.0
Hours per day watching TV	Mean	2.9
	Median	2.0
	Missing	495

Figura 43. estatísticas básicas e variáveis exibidas na dimensão da linha

Agora está claro que a tabela relata o número de valores omissos apenas para uma variável.

Agrupar sumarizações em categorias

É possível usar variáveis categóricas como variáveis de agrupamento para exibir sumarizações de variável de escala em grupos definidos pelas categorias da variável categórica.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Arraste e solte *Gênero* da lista de variáveis para a área Colunas da área de janela de tela.

Se você clicar com o botão direito em *Gênero* na visualização da tabela na área de janela de tela, verá que a opção **estatísticas básicas** está desativada no menu pop-up. Isso ocorre porque, em uma tabela com variáveis de escala, as variáveis de escala são sempre as variáveis de origem estatística.

3. Clique em **OK** para criar a tabela.

		Gender	
		Male	Female
Age of respondent	Mean	44.6	46.3
	Median	42.0	43.0
Highest year of school completed	Mean	13.4	13.2
	Median	13.0	13.0
Hours per day watching TV	Mean	2.8	2.9
	Median	2.0	2.0
	Missing	213	282

Figura 44. Sumarizações de escala agrupadas usando uma variável de coluna categórica

Essa tabela facilita a comparação das médias (média e mediana) para homens e mulheres, e mostra claramente que não há muita diferença entre eles -- o que pode não ser muito interessante, mas podem ser informações úteis.

Diversas variáveis de agrupamento

É possível subdividir os grupos ainda mais, aninhando e/ou usando variáveis de agrupamento categóricas de linha e coluna.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Arraste e solte *Obter notícias da Internet* da lista de variáveis mais à esquerda da área Linhas da área de janela de tela. Certifique-se de posicioná-la para que todas as três variáveis de escala sejam aninhadas nela, não apenas uma delas.

Embora às vezes talvez você queira algo semelhante ao segundo exemplo acima, isso não é o que queremos nesse caso.

3. Clique em **OK** para criar a tabela.

				Gender	
				Male	Female
Get news from internet	No	Age of respondent	Mean	47.0	48.8
			Median	45.0	46.0
		Highest year of school completed	Mean	13.4	13.1
			Median	13.0	12.0
		Hours per day watching TV	Mean	3.2	3.4
			Median	2.0	3.0
		Missing	213	282	
	Yes	Age of respondent	Mean	38.7	41.1
			Median	35.0	38.0
		Highest year of school completed	Mean	13.2	13.3
			Median	13.0	13.0
		Hours per day watching TV	Mean	2.1	2.1
			Median	2.0	2.0
		Missing	0	0	

Figura 45. Sumarizações de escala agrupadas por variáveis categóricas de linha e coluna

Aninhando variáveis categóricas em variáveis de escala

Embora a tabela acima possa fornecer as informações desejadas, ela pode não fornecê-las no formato mais fácil de interpretar. Por exemplo, é possível comparar a média de idade de homens que usam a Internet para obter notícias e os que não usam -- mas seria mais fácil fazer isso se os valores estivessem próximos uns dos outros em vez de separados. A troca de posições das duas variáveis de linha e o aninhamento da variável de agrupamento categórica nas três variáveis de escala podem melhorar a tabela. Com variáveis

de escala, o nível de aninhamento não tem nenhum efeito sobre a variável de origem estatística. A variável de escala é sempre a variável de origem estatística, independentemente do nível de aninhamento.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em *Idade do respondente* na visualização da tabela na área de janela de tela, Ctrl-que em *Maior ano escolar concluído*, e Ctrl-clique em *Horas por dia assistindo à TV* para selecionar todas as três variáveis de escala.
3. Arraste e solte as três variáveis de escala na extremidade esquerda da área Linhas, aninhando a variável categórica *Obter notícias da Internet* em cada uma das três variáveis de escala.
4. Clique em **OK** para criar a tabela.

				Gender	
				Male	Female
Age of respondent	Get news from internet	No	Mean	47.0	48.8
			Median	45.0	46.0
		Yes	Mean	38.7	41.1
			Median	35.0	38.0
Highest year of school completed	Get news from internet	No	Mean	13.4	13.1
			Median	13.0	12.0
		Yes	Mean	13.2	13.3
			Median	13.0	13.0
Hours per day watching TV	Get news from internet	No	Mean	3.2	3.4
			Median	2.0	3.0
			Missing	213	282
		Yes	Mean	2.1	2.1
			Median	2.0	2.0
		Missing	0	0	

Figura 46. Variável de linha categórica aninhada em variáveis de escala empilhadas

A escolha da ordem de aninhamento depende dos relacionamentos ou comparações que você deseja enfatizar na tabela. Mudar a ordem de aninhamento das variáveis de escala não altera o valor de estatísticas básicas; altera somente suas posições relativas na tabela.

Intervalos de confiança

Os intervalos de confiança e erros padrão estão disponíveis para muitas estatísticas da tabela.

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. No construtor de tabela, mova *Grau mais alto* para a área de linha da área de janela de tela.
3. Clique em **estatísticas básicas...**
4. Na lista **Estatísticas** no diálogo estatísticas básicas, clique no ícone próximo de **Contagem** para expandir a lista de estatísticas relacionadas à contagem.
5. Mova as seguintes estatísticas para a área **Exibição**: % de N da coluna, CL inferior para % de contagem de colunas, CL superior para % de contagem de colunas e Erro padrão de % de contagem de colunas.
6. No grupo **Intervalos de confiança**, para **Nível (%)**, insira 99.
7. Clique em **Aplicar à seleção** e, em seguida, clique em **Fechar**.
8. Clique em **OK** para criar a tabela.

		Count	Column N %	99.0% Lower CL for Column N %	99.0% Upper CL for Column N %	Standard Error of Column N %
Highest degree	LT High school	430	15.2%	13.6%	17.0%	0.7%
	High school	1500	53.2%	50.7%	55.6%	0.9%
	Junior college	209	7.4%	6.2%	8.7%	0.5%
	Bachelor	478	16.9%	15.2%	18.8%	0.7%
	Graduate	205	7.3%	6.1%	8.6%	0.5%

Figura 47. Tabela de contagens, porcentagens da coluna e intervalos de confiança

9. Rechame o diálogo Tabelas customizadas e clique em **estatísticas básicas...**
10. Para o limite de confiança inferior, mude o rótulo para "Limite de confiança inferior (&[Nível de confiança])". A sequência de caracteres "&[Nível de confiança]" insere o valor do nível de confiança especificado nesse local no rótulo.
11. Clique em **Aplicar à seleção** e, em seguida, clique em **Fechar**.
12. Clique em **OK** para criar a tabela.

		Count	Column N %	Lower Confidence Limit (99.0%)	99.0% Upper CL for Column N %	Standard Error of Column N %
Highest degree	LT High school	430	15.2%	13.6%	17.0%	0.7%
	High school	1500	53.2%	50.7%	55.6%	0.9%
	Junior college	209	7.4%	6.2%	8.7%	0.5%
	Bachelor	478	16.9%	15.2%	18.8%	0.7%
	Graduate	205	7.3%	6.1%	8.6%	0.5%

Figura 48. Tabela com rótulo de intervalo de confiança modificado

Estatísticas de teste

Estatísticas de teste

Três testes diferentes de significância estão disponíveis para estudar o relacionamento entre as variáveis de linha e de coluna. Este capítulo discute a saída de cada um desses testes, com especial atenção aos efeitos do aninhamento e do empilhamento. Consulte o tópico "Empilhamento, aninhamento e estratos com variáveis categóricas" na página 25 para obter mais informações

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *survey_sample.sav*. Consulte o tópico arquivos de dados para obter mais informações

Testes de independência (Qui-quadrado)

O teste qui-quadrado de independência é usado para determinar se há um relacionamento entre duas variáveis categóricas. Por exemplo, talvez você queira determinar se *Status da força de mão de obra* está relacionado a *Estado civil*.

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. No construtor de tabela, arraste e solte *Status da força de mão de obra* da lista de variáveis para a área Linhas da área de janela de tela.
3. Arraste e solte *Estado civil* da lista de variáveis para a área Colunas.

4. Selecione **Linhas** como a posição para as estatísticas básicas.
5. Selecione *Status da força de mão de obra* e clique em **estatísticas básicas** no grupo Definição.
6. Selecione **% de N da coluna** na lista Estatísticas e inclua-a na lista Exibição.
7. Clique em **Aplicar à seleção**.
8. Na caixa de diálogo Tabelas customizadas, clique na guia **Estatísticas de teste**.
9. Selecione **Testes de independência (Qui-quadrado)**.
10. Clique em **OK** para criar a tabela e obter o teste qui-quadrado.

			Marital status				
			Married	Widowed	Divorced	Separated	Never married
Labor force status	Working full time	Count	778	44	295	58	392
		Column %	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	Count	138	20	35	9	102
		Column %	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	Count	23	2	9	1	11
		Column %	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	Count	13	3	10	0	32
		Column %	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	Count	168	150	53	6	17
		Column %	12.5%	53.0%	11.9%	6.5%	2.6%
	School	Count	9	1	7	2	60
		Column %	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	Count	200	55	25	13	35
		Column %	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	Count	16	8	12	4	14
		Column %	1.2%	2.8%	2.7%	4.3%	2.1%

Figura 49. Status da força de mão de obra por Estado civil

Essa tabela é uma tabulação cruzada de *Status da força de mão de obra* por *Estado civil*, com contagens e proporções da coluna mostradas como estatísticas básicas. As proporções da coluna são calculadas para que somem 100% em cada coluna. Se essas duas variáveis não estiverem relacionadas, em cada linha, as proporções devem ser semelhantes entre as colunas. Parece haver diferenças nas proporções, mas é possível verificar o teste qui-quadrado para ter certeza.

		Marital status
Labor force status	Chi-square	729.242
	df	28
	Sig.	.000*

*. The Chi-square statistic is significant at the 0.05 level.

Figura 50. Teste qui-quadrado de Pearson

O teste de independência levanta uma hipótese de que o *Status da força de mão de obra* e *Estado civil* não estão relacionados -- ou seja, que as proporções da coluna são iguais entre as colunas, e as discrepâncias observadas são devido à variação de chances. A estatística qui-quadrado mede a discrepância geral entre as contagens de células observadas e as contagens que seriam esperadas se as proporções da coluna fossem iguais entre as colunas. Uma estatística qui-quadrado maior indica uma maior discrepância entre as contagens de células observadas e esperadas -- evidência maior de que as proporções da coluna não são iguais, que a hipótese de independência está incorreta e, portanto, que *Status da força de mão de obra* e *Estado civil* estão relacionados.

A estatística qui-quadrado calculada tem um valor de 729.242. Para determinar se isso é evidência suficiente para rejeitar a hipótese de independência, o valor de significância da estatística é calculado. O valor de significância é a probabilidade de que uma variável aleatória obtida de uma distribuição

qui-quadrado com 28 graus de liberdade é maior que 729.242. Como esse valor é menor que o nível alpha especificado na guia Estatísticas de teste, é possível rejeitar a hipótese de independência no nível 0,05. Portanto, *Status da força de mão de obra* e *Estado civil* estão de fato relacionados.

Efeitos do aninhamento e do empilhamento em testes de independência

A regra para testes de independência é a seguinte: é executado um teste separado para cada subtabela mais interna. Para ver como o aninhamento afeta os testes, considere o exemplo anterior, mas com *Estado civil* aninhado nos níveis de *Gênero*.

1. Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
2. Arraste e solte *Gênero* da lista de variáveis para a área Colunas da área de janela de tela, acima de *Estado civil*.
3. Clique em **OK** para criar a tabela.

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	Chi-square	246.637	542.589
	df	28	28
	Sig.	.000 ^a .1,2	.000 ^a .1,2

*. The Chi-square statistic is significant at the 0.05 level.

1. More than 20% of cells in this sub-table have expected cell counts less than 5.
2. The minimum expected cell count in this sub-table is less than one.

Figura 51. Teste qui-quadrado de Pearson

Com *Estado civil* aninhado em níveis de *Gênero*, são executados dois testes -- um para cada nível de *Gênero*. O valor de significância para cada teste indica que é possível rejeitar a hipótese de independência entre *Estado civil* e *Status da força de mão de obra* para homens e mulheres. No entanto, a tabela nota que mais de 20% das células de cada tabela possuem contagens esperadas menores que 5, e a contagem mínima esperada de células é menor que 1. Essas notas indicam que as suposições do teste qui-quadrado podem não ser atendidas por essas tabelas, portanto, os resultados dos testes são suspeitos.

Nota: As notas de rodapé podem ser cortadas da visualização pelos limites da célula. É possível torná-las visíveis mudando o alinhamento dessas células na caixa de diálogo Propriedades da célula.

Para ver como o empilhamento afeta os testes:

4. Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
5. Arraste e solte *Grau mais alto* da lista de variáveis para a área Linhas, abaixo de *Status da força de mão de obra*.
6. Clique em **OK** para criar a tabela.

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	Chi-square	246.637	542.589
	df	28	28
	Sig.	.000 ^a .1,2	.000 ^a .1,2
Highest degree	Chi-square	43.844	105.506
	df	16	16
	Sig.	.000 ^a	.000 ^a

*. The Chi-square statistic is significant at the 0.05 level.

1. More than 20% of cells in this sub-table have expected cell counts less than 5.
2. The minimum expected cell count in this sub-table is less than one.

Figura 52. Teste qui-quadrado de Pearson

Com *Grau mais alto* empilhado com *Status da força de mão de obra*, são executados quatro testes -- um teste da independência de *Estado civil* e *Status da força de mão de obra* e um teste de *Estado civil* e *Grau mais alto* para cada nível de *Gênero*. Os resultados do teste para *Estado civil* e *Status da força de mão de obra* são os mesmos de antes. Os resultados do teste para *Estado civil* e *Grau mais alto* indicam que essas variáveis não são independentes.

Comparando médias de coluna

Os testes de médias de coluna são usados para determinar se há um relacionamento entre uma variável categórica nas Colunas e uma variável contínua nas Linhas. Além disso, é possível usar os resultados do teste para determinar a ordem relativa de categorias da variável categórica em termos do valor médio da variável contínua. Por exemplo, talvez você queira determinar se *Horas por dia assistindo à TV* está relacionado a *Obter notícias de jornais*.

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. Clique em **Reconfigurar** para restaurar as configurações padrão para todas as guias.
3. No construtor de tabela, arraste e solte *Horas por dia assistindo à TV* da lista de variáveis para a área Linhas da área de janela de tela.
4. Arraste e solte *Obter notícias de jornais* da lista de variáveis para a área Colunas.
5. Selecione *Horas por dia assistindo à TV* e clique em **estatísticas básicas** no grupo Definição.
6. Selecione **nnnn** como o formato.
7. Selecione **2** como o número de casas decimais a serem exibidas. Observe que isso agora faz o formato ser lido como **nnnn.nn**.
8. Clique em **Aplicar à seleção**.
9. Na caixa de diálogo Tabelas customizadas, clique na guia **Estatísticas de teste**.
10. Selecione **Comparar médias de coluna (testes t)**.
11. Clique em **OK** para criar a tabela e obter os testes de médias de coluna.

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92	2.74

Figura 53. Obter notícias de jornais por Horas por dia assistindo à TV

Essa tabela mostra a média *Horas por dia assistindo à TV* para pessoas que obtêm e não obtêm suas notícias de jornais. A diferença observada nessas médias sugere que as pessoas que não obtêm suas notícias de jornais gastam aproximadamente 0,18 mais horas assistindo à TV do que as pessoas que obtêm suas notícias de jornais. Para verificar se essa diferença é devido à variação de chances, verifique os testes de médias de coluna.

	Get news from newspapers	
	No	Yes
	(A)	(B)
Hours per day watching TV		

Figura 54. Comparações de médias de coluna

A tabela de teste de médias de coluna designa uma chave de letra a cada categoria da variável da coluna. Para *Obter notícias de jornais*, a categoria *Não* é designada à letra A, e *Sim* é designado à letra B. Para cada par de colunas, as médias de coluna são comparadas usando um teste *t*. Como há apenas duas colunas, apenas um teste é executado. Para cada par significativo, a chave da categoria com a menor média é colocada na categoria com a maior média. Como não há chaves relacionadas nas células da tabela, isso significa que as médias de coluna não são estatisticamente diferentes.

Resultados de significância na notação de estilo APA

Se não desejar os resultados de significância em uma tabela separada, será possível optar por exibi-los na tabela principal. Os resultados de significância são identificados usando uma notação de estilo APA com letras subscritas. Conclua os passos anteriores para comparar médias de coluna, mas faça a seguinte mudança na guia Estatísticas de teste:

1. Na área Identificar diferenças significativas, selecione **Na tabela principal usando subscritos de estilo APA**.
2. Clique em **OK** para criar a tabela e obter os testes de médias de coluna usando a notação de estilo APA.

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92 _a	2.74 _a

Figura 55. Comparações de médias de coluna usando a notação de estilo APA

A tabela de teste de médias de coluna designa uma letra subscrita às categorias da variável da coluna. Para cada par de colunas, as médias de coluna são comparadas usando um teste *t*. Se um par de valores for significativamente diferente, os valores terão *diferentes* letras subscritas designadas a eles. Como há apenas duas colunas, apenas um teste é executado. Como as médias de coluna nesse exemplo compartilham a mesma letra subscrita, as médias de coluna não são estatisticamente diferentes.

Efeitos do aninhamento e do empilhamento em testes de médias de coluna

A regra para testes de médias de coluna é a seguinte: um conjunto separado de testes de pares é executado para cada subtabela mais interna. Para ver como o aninhamento afeta os testes, considere o exemplo anterior, mas com *Horas por dia assistindo à TV* aninhado em níveis de *Status da força de mão de obra*.

1. Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
2. Arraste e solte *Status da força de mão de obra* da lista de variáveis para a área Linhas da área de janela de tela.
3. Clique em **OK** para criar a tabela.

			Get news from newspapers	
			No	Yes
			(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B	
	Working part-time	Hours per day watching TV		
	Temporarily not working	Hours per day watching TV		
	Unemployed, laid off	Hours per day watching TV		
	Retired	Hours per day watching TV		
	School	Hours per day watching TV		
	Keeping house	Hours per day watching TV		
	Other	Hours per day watching TV		

Figura 56. Comparações de médias de coluna

Com *Horas por dia assistindo à TV* aninhado nos níveis de *Status da força de mão de obra*, sete conjuntos de testes de médias de coluna são executados: um para cada nível de *Status da força de mão de obra*. As mesmas chaves de letra são designadas às categorias de *Obter notícias de jornais*. Para respondentes *trabalhando em período integral*, a chave B aparece na coluna A. Isso significa que, para funcionário de período integral, o valor médio de *Horas por dia assistindo à TV* é inferior para pessoas que obtêm suas notícias de jornais. Nenhuma outra chave aparece nas colunas, portanto, é possível concluir que não há outras diferenças estatisticamente significativas nas médias de coluna.

Correções de Bonferroni. Quando vários testes são executados, a correção de Bonferroni é aplicada a testes de médias de coluna para assegurar que o nível alfa (ou taxa positiva falsa) especificado na guia Estatísticas de teste se aplicará a cada *conjunto* de testes. Portanto, nessa tabela, nenhuma correção de Bonferroni foi aplicada porque, embora sete conjuntos de testes tenham sido executados, em cada conjunto, apenas um par de colunas é comparado.

Para ver como o empilhamento afeta os testes:

4. Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
5. Arraste e solte *Obter notícias da Internet* da lista de variáveis para a área Colunas, à esquerda de *Obter notícias de jornais*.
6. Clique em **OK** para criar a tabela.

			Get news from internet		Get news from newspapers	
			No	Yes	No	Yes
			(A)	(B)	(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B		B	
	Working part-time	Hours per day watching TV	B			
	Temporarily not working	Hours per day watching TV				
	Unemployed, laid off	Hours per day watching TV	B			
	Retired	Hours per day watching TV	B			
	School	Hours per day watching TV	B			
	Keeping house	Hours per day watching TV	B			
	Other	Hours per day watching TV	B			

Figura 57. Comparações de médias de coluna

Com *Obter notícias da Internet* empilhado com *Obter notícias de jornais*, 14 conjuntos de testes de médias de coluna são executados -- um para cada nível de *Status da força de mão de obra* para *Obter notícias da Internet* e *Obter notícias de jornais*. Novamente, nenhuma correção de Bonferroni foi aplicada, porque, em cada conjunto, apenas um par de colunas é comparado. Os testes para *Obter notícias de jornais* são os mesmos de antes. Para *Obter notícias da Internet*, a categoria *Não* é designada à letra A e *Sim* é designada à letra B. A chave B é relatada na coluna A para cada conjunto de testes de médias de coluna, exceto para os respondentes que não estão trabalhando temporariamente. Isso significa que o valor médio de *Horas por dia assistindo à TV* é inferior para pessoas que obtêm suas notícias da Internet do que para pessoas que não obtêm suas notícias de jornais. Nenhuma chave é relatada para o conjunto *Não trabalhando temporariamente*; portanto, as médias de coluna não são estatisticamente diferentes para esses respondentes.

Comparando proporções da coluna

Os testes de proporções da coluna são usados para determinar a ordem relativa de categorias da variável categórica de Colunas em termos das proporções de categoria da variável categórica de Linhas. Por exemplo, depois de usar um teste qui-quadrado para saber que *Status da força de mão de obra* e *Estado civil* não são independentes, é possível saber quais linhas e colunas são responsáveis por esse relacionamento.

1. Nos menus, escolha:
Analisar > Tabelas > Tabelas Customizadas...
2. Clique em **Reconfigurar** para restaurar as configurações padrão para todas as guias.
3. No construtor de tabela, arraste e solte *Status da força de mão de obra* da lista de variáveis para a área Linhas da área de janela de tela.
4. Arraste e solte *Estado civil* da lista de variáveis para a área Colunas.
5. Selecione *Status da força de mão de obra* e clique em **estatísticas básicas** no grupo Definição.
6. Selecione **% de N da coluna** na lista Estatísticas e inclua-a na lista Exibição.
7. Cancele a seleção de **Contagem** na lista Exibição.
8. Clique em **Aplicar à seleção**.
9. Na caixa de diálogo Tabelas customizadas, clique na guia **Estatísticas de teste**.
10. Selecione **Comparar proporções da coluna (testes z)**.
11. Clique em **OK** para criar a tabela e obter os testes de proporções da coluna.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Column %	Column %	Column %	Column %	Column %
Labor force status	Working full time	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	12.5%	53.0%	11.9%	6.5%	2.6%
	School	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Figura 58. Status da força de mão de obra por Estado civil

Essa tabela é uma tabulação cruzada de *Status da força de mão de obra por Estado civil*, com proporções da coluna mostradas como a estatística de sumarização.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(A)	(B)	(C)	(D)	(E)
Labor force status	Working full time	B		A B	B	B
	Working part-time					A B C
	Temporarily not working					
	Unemployed, laid off					A B
	Retired	E	A C D E	E		
	School					A B C
	Keeping house	C E	C E		C E	
	Other					

Figura 59. Comparações de proporções da coluna

A tabela de teste de proporções da coluna designa uma chave de letra a cada categoria das variáveis da coluna. Para *Estado civil*, a categoria *Casado* é designada à letra A, *Viúvo* é designada à letra B, e assim por diante, até a categoria *Nunca se casou*, que é designada à letra E. Para cada par de colunas, as proporções da coluna são comparadas usando um teste z. São executados sete conjuntos de proporções da coluna, um para cada nível de *Status de força de mão de obra*. Como existem cinco níveis de *Estado civil*, $(5^*4)/2 = 10$ pares de colunas são comparados em cada conjunto de testes, e as correções de Bonferroni são usadas para ajustar os valores de significância. Para cada par significativo, a chave da categoria menor é colocada na categoria com a maior proporção.

Para o conjunto de testes associados a *Trabalhando em período integral*, a chave B aparece em cada uma das outras colunas. Além disso, a chave A aparece na coluna C. Nenhuma outra chave é relatada em outras colunas. É possível concluir que a proporção de pessoas divorciadas que estão trabalhando em período integral é maior que a proporção de pessoas casadas que trabalham em período integral, que, por sua vez, é maior que a proporção de viúvos que trabalham em período integral. As proporções de pessoas que são separadas ou nunca se casaram e trabalham em período integral não podem ser diferenciadas de pessoas divorciadas ou casadas e trabalham em período integral, mas essas proporções são maiores que a proporção de viúvos que trabalham em período integral.

Para os testes associados a *Trabalhando em período integral* ou *Escola*, as chaves A, B e C aparecem na coluna E. Nenhuma outra chave é relatada em outras colunas. Portanto, as proporções de pessoas que

nunca foram casadas e estão na escola ou que trabalham meio período são maiores que as proporções de pessoas casadas, viúvas ou divorciadas que estão na escola ou trabalham meio período.

Para os testes associados a *Não trabalhando temporariamente* ou a *Outro* status de mão de obra, nenhuma outra chave é relatada em nenhuma coluna. Portanto, não há nenhuma diferença perceptível nas proporções de pessoas casadas, viúvas, divorciadas, separadas ou que nunca se casaram que não estão trabalhando temporariamente ou estão, de outra forma, em uma situação de emprego não categorizada.

Os testes associados a *Aposentado* mostram que a proporção de viúvos que são aposentados é maior que as proporções de todas as outras categorias de estado civil de pessoas que são aposentadas. Além disso, as proporções de pessoas casadas ou divorciadas que são aposentadas é maior que a proporção de pessoas que nunca se casaram e são aposentadas.

Há maiores proporções de pessoas casadas, viúvas ou separadas e que cuidam da casa do que proporções de pessoas divorciadas ou que nunca se casaram e cuidam da casa.

A proporção de pessoas que nunca se casaram e estão *Desempregadas, demitidas* é mais alta do que as proporções de pessoas que são casadas ou viúvas e desempregadas. Além disso, observe que a coluna *Separado* está marcada com um ".", que indica que a proporção observada de pessoas separadas na linha *Desempregadas, demitidas* é 0 ou 1 e, portanto, nenhuma comparação pode ser feita usando essa coluna para respondentes desempregados.

Mesclando resultados de significância na tabela principal

Se não desejar os resultados de significância em uma tabela separada, será possível optar por exibi-los na tabela principal. Conclua os passos anteriores para comparar proporções da coluna, mas faça a seguinte mudança na guia Estatísticas de teste:

1. Na área Identificar diferenças significativas, selecione **Na tabela principal**.
2. Clique em **OK** para criar a tabela.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(a)	(b)	(c)	(d)	(e)
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% b	15.5%	66.1% a b	62.4% b	59.1% b
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4% a b c
	Temporarily not working	1.7%	0.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	0.0% ¹	4.8% a b
	Retired	12.5% e	53.0% a c d e	11.9% e	6.5%	2.6%
	School	0.7%	0.4%	1.6%	2.2%	9.0% a b c
	Keeping house	14.9% c e	19.4% c e	5.6%	14.0% c e	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Figura 60. Resultados de significância mesclados na tabela principal

3. Clique duas vezes na tabela na janela Visualizador para ativá-la.

4. Passe o mouse sobre uma das células de rótulos da coluna que contém as chaves de letras. Por exemplo, passe o mouse sobre a célula com o rótulo "(a)".

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(a)	(b)	(c)	(d)	(e)
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% b	15.5%	66.1% a b	62.4% b	59.1% b
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4% a b c
	Temporarily not working	1.7%	0.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	0.0% ¹	4.8% a b
	Retired	12.5% e	53.0% a c d e	11.9% e	6.5%	2.6%
	School	0.7%	0.4%	1.6%	2.2%	9.0% a b c
	Keeping house	14.9% c e	19.4% c e	5.6%	14.0% c e	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Figura 61. Destacando todas as células que contêm a chave de letra selecionada

Todas as células que contêm essa chave são destacadas na tabela.

5. Clique com o botão direito na célula do rótulo da coluna com a chave de letra. No menu de contexto, escolha **Selecionar > Selecionar todas as células com essa chave de significância**.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(a)	(b)	(c)	(d)	(e)
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% b	15.5%	66.1% a b	62.4% b	59.1% b
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4% a b c
	Temporarily not working	1.7%	0.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	0.0% ¹	4.8% a b
	Retired	12.5% e	53.0% a c d e	11.9% e	6.5%	2.6%
	School	0.7%	0.4%	1.6%	2.2%	9.0% a b c
	Keeping house	14.9% c e	19.4% c e	5.6%	14.0% c e	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Figura 62. Selecionando todas as células que contêm a chave de letra selecionada

Todas as células que contêm essa chave são selecionadas na tabela.

Efeitos do aninhamento e empilhamento em testes de proporções da coluna

A regra para testes de proporções da coluna é a seguinte: um conjunto separado de testes de pares é executado para cada subtabela mais interna. Para saber como o aninhamento afeta os testes, considere o exemplo anterior, mas com *Status da força de mão de obra* aninhado em níveis de *Gênero*.

1. Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
2. Arraste e solte *Gênero* da lista de variáveis para a área Linhas da área de janela de tela.
3. Clique em **OK** para criar a tabela.

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working		.			
			Unemployed, laid off		.		.	A
			Retired	E	A C D E	E		
			School		.			A C
			Keeping house					
	Other				A			
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working				.	
			Unemployed, laid off				.	A
			Retired	E	A C D E	E		
			School					A B C
Keeping house			C E	C E		C		
Other								

Figura 63. Comparações de proporções da coluna

Com *Status da força de mão de obra* aninhado em níveis de *Gênero*, são executados 14 conjuntos de testes de proporções da coluna -- um para cada nível de *Status da força de mão de obra* para cada nível de *Gênero*. As mesmas chaves de letra são designadas às categorias de *Estado civil*.

Há algumas coisas a serem observadas sobre os resultados da tabela:

- Com mais testes, há mais colunas com zero proporção da coluna. Eles são mais comuns entre respondentes separados e homens viúvos.
- As diferenças das colunas vistas anteriormente entre respondentes que *cuidam da casa* parece ser totalmente por causa das mulheres.

Para ver como o empilhamento afeta os testes:

4. Abra o construtor de tabela novamente (menu Analisar, Tabelas, Tabelas customizadas).
5. Arraste e solte *Grau mais alto* da lista de variáveis para a área Linhas, abaixo de *Gênero*.
6. Clique em **OK** para criar a tabela.

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A C
			Keeping house					
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A B C
			Keeping house	C E	C E		C	
Highest degree		Other		A C E				
			LT High school					
			High school					
			Junior college	B		B		B
			Bachelor	B				B
			Graduate	B				

Figura 64. Comparações de proporções da coluna

Com *Grau mais alto* empilhado com *Gênero*, são executados 19 conjuntos de testes de médias de coluna -- os 14 discutidos anteriormente, além de um para cada nível de *Grau mais alto*. As mesmas chaves de letra são designadas às categorias de *Estado civil*.

Há algumas coisas a serem observadas sobre os resultados da tabela:

- Os resultados do teste para os 14 conjuntos de testes executados anteriormente são os mesmos.
- As pessoas com um grau menor que o ensino secundário são mais comuns entre viúvos do que entre respondentes casados, divorciados ou que nunca se casaram.
- As pessoas com um nível de educação superior ao ensino secundário tendem a ser mais comuns entre as pessoas que são casadas, divorciadas e que nunca se casaram do que entre viúvos.

Uma nota sobre ponderações e conjuntos de múltiplas respostas

As ponderações de caso são sempre baseadas em contagens, não em respostas, mesmo quando uma das variáveis é uma variável de respostas diversas.

Conjuntos de múltiplas respostas

Tabelas Customizadas e o Construtor de Gráfico suportam um tipo especial de "variável" chamado **conjunto de múltiplas respostas**. Os conjuntos de múltiplas respostas não são na realidade "variáveis" no sentido normal. Não é possível vê-los no Editor de Dados, e outros procedimentos não os reconhecem. Os conjuntos de múltiplas respostas utilizam diversas variáveis para registrar respostas para perguntas em que o respondente pode fornecer mais de uma resposta. Os conjuntos de múltiplas respostas são tratados como variáveis categóricas e a maioria das coisas que podem ser feitas com variáveis categóricas também podem ser feitas com conjuntos de múltiplas respostas.

Os conjuntos de múltiplas respostas são construídos a partir de diversas variáveis no arquivo de dados. Um conjunto de múltiplas respostas é uma construção especial dentro de um arquivo de dados. É possível definir e salvar conjuntos de múltiplas respostas em arquivos de dados IBM® SPSS Statistics, mas não é possível importar ou exportar conjuntos de múltiplas respostas de/para outros formatos de arquivo. É possível copiar conjuntos de múltiplas respostas a partir de outros arquivos de dados do IBM SPSS Statistics utilizando Copiar Propriedades de Dados, que é acessado a partir do menu Dados na janela Editor de Dados.

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *survey_sample.sav*. Consulte o tópico arquivos de dados para obter mais informações

Todos os exemplos fornecidos aqui exibem rótulos de variáveis em caixas de diálogo, ordenados em ordem alfabética. As propriedades de exibição da lista de variáveis são especificadas na guia Geral na caixa de diálogo Opções (menu Editar, Opções).

Contagens, respostas, porcentagens e totais

Todas as estatísticas básicas disponíveis para variáveis categóricas também estão disponíveis para conjuntos de múltiplas respostas. Algumas estatísticas adicionais também estão disponíveis para conjuntos de múltiplas respostas.

1. Nos menus, escolha:

Analisar > Tabelas > Tabelas Customizadas...

2. Arraste e solte *Fontes de notícias* (esse é o rótulo descritivo para o conjunto de múltiplas respostas *\$mltnews*) da lista de variáveis para a área Linhas da área de janela de tela.

O ícone próximo à "variável" na lista de variáveis identifica-a como um conjunto de múltiplas dicotomias.



Figura 65. Ícone do conjunto de múltiplas dicotomias

Para um conjunto de múltiplas dicotomias, cada "categoria" é, na verdade, uma variável separada, e os rótulos de categoria são os rótulos de variáveis (ou nomes de variáveis para variáveis sem rótulos de variáveis definidos). Nesse exemplo, as contagens que serão exibidas representam o número de casos com uma resposta *Sim* para cada variável no conjunto.

3. Clique com o botão direito em *Fontes de notícias* na visualização da tabela na área de janela de tela e selecione **Categorias e totais** no menu pop-up.
4. Selecione (clique) **Total** na caixa de diálogo Categorias e totais e, em seguida, clique em **Aplicar**.
5. Clique com o botão direito em *Fontes de notícias* novamente e selecione **estatísticas básicas** no menu pop-up.
6. Na caixa de diálogo estatísticas básicas, selecione **% de N da coluna** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
7. Clique em **Aplicar à seleção** e, em seguida, clique em **OK** para criar a tabela.

		Count	Column N %
News sources	Get news from internet	867	41.7%
	Get news from radio	551	26.5%
	Get news from television	1077	51.8%
	Get news from news magazines	294	14.1%
	Get news from newspapers	805	38.7%
	Total	2081	100.0%

Figura 66. Contagens de dicotomias múltiplas e porcentagens da coluna

Totais que não correspondem

Se você examinar os números na tabela, poderá observar que há uma grande discrepância entre os "totais" e os valores que supostamente estão sendo totalizados -- especificamente, os totais parecem ser muito menores do que deveriam ser. Isso ocorre porque a contagem para cada "categoria" na tabela é o número de casos com um valor 1 (uma resposta *Sim*) para essa variável, e o número total de respostas *Sim* para todas as cinco variáveis no conjunto de múltiplas dicotomias pode facilmente exceder o número total de casos no arquivo de dados.

No entanto, a "contagem" total é o número total de casos com uma resposta *Sim* para pelo menos uma variável no conjunto, que nunca pode exceder o número total de casos no arquivo de dados. Neste exemplo, a contagem total de 2.081 é quase 800 mais baixa do que o número total de casos no arquivo de dados. Se nenhuma dessas variáveis tiver valores omissos, isso significa que quase 800 respondentes da pesquisa de opinião indicaram que não recebem notícias de nenhuma dessas fontes. A contagem total é a base para as porcentagens da coluna; portanto, as porcentagens da coluna nesse exemplo somam mais que os 100% exibidos para o total da porcentagem da coluna.

Totais que correspondem

Embora "contagem" geralmente seja um termo bastante inequívoco, o exemplo acima demonstra como ele pode ser confuso no contexto de totais para conjuntos de múltiplas respostas, para os quais as *respostas* geralmente são a estatística de sumarização realmente desejada.

1. Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas).
2. Clique com o botão direito em *Fontes de notícias* na visualização da tabela na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
3. Na caixa de diálogo estatísticas básicas, selecione **Respostas** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
4. Selecione **% de respostas da coluna** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
5. Clique em **Aplicar à seleção** e, em seguida, clique em **OK** para criar a tabela.

		Count	Column N %	Responses	Column Responses %
News sources	Get news from internet	867	41.7%	867	24.1%
	Get news from radio	551	26.5%	551	15.3%
	Get news from television	1077	51.8%	1077	30.0%
	Get news from news magazines	294	14.1%	294	8.2%
	Get news from newspapers	805	38.7%	805	22.4%
	Total	2081	100.0%	3594	100.0%

Figura 67. Respostas de múltiplas dicotomias e porcentagens de respostas da coluna

Para cada "categoria" no conjunto de múltiplas dicotomias, as *Respostas* são idênticas à *Contagem* -- e esse sempre será o caso para conjuntos de múltiplas dicotomias. No entanto, os totais são muito diferentes. O número total de respostas é de 3.594 -- mais de 1.500 maior que a contagem total e mais de 700 maior que o número total de casos no arquivo de dados.

Para porcentagens, os totais para % de N da coluna e % de respostas da coluna são de 100% -- mas as porcentagens para cada categoria no conjunto de múltiplas dicotomias são muito menores para porcentagens de respostas da coluna. Isso ocorre porque a base de porcentagem para porcentagens de respostas da coluna é o número total de respostas que, nesse caso, é 3.594, resultando em porcentagens muito menores que a base de porcentagem da coluna de 2.081.

Porcentagens totais maiores que 100%

As porcentagens da coluna e as porcentagens de respostas da coluna resultam em porcentagens totais de 100% mesmo que, em nosso exemplo, os valores individuais na coluna % N da coluna somem mais de 100%. Mas, e se você desejar mostrar porcentagens com base na contagem total em vez do total de respostas, mas também desejar que a porcentagem "total" reflita de forma precisa a soma das porcentagens de categorias individuais?

1. Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas).
2. Clique com o botão direito em *Fontes de notícias* na visualização da tabela na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
3. Na caixa de diálogo estatísticas básicas, selecione **% de respostas da coluna (Base: Contagem)** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
4. Clique em **Aplicar à seleção** e, em seguida, clique em **OK** para criar a tabela.

		Count	Column N %	Responses	Column Responses %	Column Responses % (Base: Count)
News sources	Get news from internet	867	41.7%	867	24.1%	41.7%
	Get news from radio	551	26.5%	551	15.3%	26.5%
	Get news from television	1077	51.8%	1077	30.0%	51.8%
	Get news from news magazines	294	14.1%	294	8.2%	14.1%
	Get news from newspapers	805	38.7%	805	22.4%	38.7%
Total		2081	100.0%	3594	100.0%	172.7%

Figura 68. Porcentagens de respostas da coluna com contagem como a base de porcentagem

Usando conjuntos de múltiplas respostas com outras variáveis

Em geral, é possível usar conjuntos de múltiplas respostas assim como variáveis categóricas. Por exemplo, é possível fazer a tabulação cruzada de um conjunto de múltiplas respostas com uma variável categórica ou aninhar um conjunto de múltiplas respostas em uma variável categórica.

1. Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas).
2. Arraste e solte *Gênero* da lista de variáveis do lado esquerdo da área Linhas na área de janela de visualização, aninhando as *Origens de notícias* do conjunto de múltiplas respostas em categorias de gênero.
3. Clique com o botão direito em *Gênero* na visualização da tabela na área de janela de tela e cancele a seleção de **Mostrar rótulo de variáveis** no menu pop-up.
4. Faça o mesmo para *Origens de notícias*.
Isso removerá as colunas com os rótulos de variáveis da tabela (pois elas de fato não são necessárias nesse caso).
5. Clique em **OK** para criar a tabela.

		Count	Column N %	Responses	Column Responses %	Column Responses % (Base: Count)
Male	Get news from internet	359	40.1%	359	23.3%	40.1%
	Get news from radio	233	26.0%	233	15.1%	26.0%
	Get news from television	451	50.3%	451	29.3%	50.3%
	Get news from news magazines	121	13.5%	121	7.9%	13.5%
	Get news from newspapers	375	41.9%	375	24.4%	41.9%
	Total	896	100.0%	1539	100.0%	171.8%
Female	Get news from internet	508	42.9%	508	24.7%	42.9%
	Get news from radio	318	26.8%	318	15.5%	26.8%
	Get news from television	626	52.8%	626	30.5%	52.8%
	Get news from news magazines	173	14.6%	173	8.4%	14.6%
	Get news from newspapers	430	36.3%	430	20.9%	36.3%
	Total	1185	100.0%	2055	100.0%	173.4%

Figura 69. Conjunto de múltiplas respostas aninhado em uma variável categórica

Variável de origem estatística e estatísticas básicas disponíveis

Na ausência de uma variável de escala em uma tabela, as variáveis categóricas e conjuntos de múltiplas respostas são tratados da mesma forma com relação à variável de origem estatística: A variável aninhada mais interna na dimensão de origem estatística é a variável de origem estatística. Como há algumas estatísticas básicas que podem ser designadas apenas a conjuntos de múltiplas respostas, isso significa que o conjunto de múltiplas respostas deve ser a variável aninhada mais interna na dimensão de origem estatística, se você desejar alguma das estatísticas básicas de múltiplas respostas especiais.

1. Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas).
2. Na visualização da tabela na área de janela de tela, arraste e solte *Fontes de notícias* à esquerda de *Gênero*, mudando a ordem de aninhamento.

Todas as estatísticas básicas especiais de múltiplas respostas -- respostas, porcentagens de respostas da coluna -- são removidas da visualização da tabela porque a variável categórica *Gênero* agora é a variável aninhada mais interna e, portanto, a variável de origem estatística.

Felizmente, o construtor de tabela "lembra" essas configurações. Se você mover *Fontes de notícias* de volta para sua posição anterior, aninhado em *Gênero*, todas as estatísticas básicas relacionadas à resposta serão restauradas para a visualização da tabela.

Conjuntos de categorias múltiplas e respostas duplicadas

Os conjuntos de categorias múltiplas fornecem um recurso não disponível para conjuntos de múltiplas dicotomias: a capacidade de contar respostas duplicadas. Em muitos casos, as respostas duplicadas em conjuntos de categorias múltiplas representam erros de codificação. Por exemplo, para uma pergunta de pesquisa de opinião, como "Em sua opinião, quais são os três países que fabricam os melhores carros?", uma resposta de *Suécia, Alemanha e Suécia* provavelmente não é válida.

No entanto, em outros casos, respostas duplicadas podem ser perfeitamente válidas. Por exemplo, se a pergunta fosse: "Onde foram fabricados seus três últimos carros?", uma resposta de *Suécia, Alemanha e Suécia* faz todo o sentido.

As Tabelas customizadas fornecem uma opção para respostas duplicadas em conjuntos de categorias múltiplas. Por padrão, as respostas duplicadas não são contadas, mas é possível solicitar que elas sejam incluídas.

1. Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas).
2. Clique em **Reconfigurar** para limpar as configurações anteriores.
3. Arraste e solte *Fabricante de automóveis, automóveis mais recentes* da lista de variáveis para a área Linhas da área de janela de tela.

O ícone próximo à "variável" na lista de variáveis identifica-a como um conjunto de categorias múltiplas.



Figura 70. Ícone Conjunto de categorias múltiplas

Para conjuntos de categorias múltiplas, as categorias exibidas representam o conjunto comum de rótulos de valor definidos para todas as variáveis no conjunto (enquanto para conjuntos de múltiplas dicotomias, as "categorias" são, na verdade, os rótulos de variável para cada variável no conjunto).

4. Clique com o botão direito em *Fabricante de automóveis, automóveis mais recentes* na visualização da tabela, na área de janela de tela, e selecione **Categorias e totais** no menu pop-up.
5. Selecione (clique) **Total** na caixa de diálogo Categorias e totais e, em seguida, clique em **Aplicar**.
6. Clique com o botão direito em *Fabricante de automóveis, automóveis mais recentes* novamente e selecione **estatísticas básicas** no menu pop-up.
7. Na caixa de diálogo estatísticas básicas, selecione **Respostas** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
8. Clique em **Aplicar à seleção** e, em seguida, clique em **OK** para criar a tabela.

		Count	Responses
Car maker, most recent cars	American	1938	1938
	Japanese	1327	1327
	Korean	695	695
	German	693	693
	Swedish	360	360
	Other	343	343
	Total	2832	5356

Figura 71. Conjunto de categorias múltiplas: Contagens e respostas sem duplicatas

Por padrão, as respostas duplicadas não são contadas; portanto, nesta tabela, os valores para cada categoria nas colunas *Contagem* e *Respostas* são idênticos. Apenas os totais diferem.

9. Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas).
10. Clique na guia **Opções**.
11. Clique em (marque) **Contar respostas duplicadas para conjuntos de categorias múltiplas**.
12. Clique em **OK** para criar a tabela.

		Count	Responses
Car maker, most recent cars	American	1938	2797
	Japanese	1327	1717
	Korean	695	760
	German	693	754
	Swedish	360	383
	Other	343	359
	Total	2832	6770

Figura 72. Conjunto de categorias múltiplas com respostas duplicadas incluídas

Nessa tabela, há uma diferença muito notável entre os valores nas colunas *Contagem* e *Respostas*, principalmente para carros norte-americanos, indicando que muitos respondentes tiveram vários carros norte-americanos.

Teste de significância com conjuntos de múltiplas respostas

É possível usar conjuntos de múltiplas respostas em testes de significância basicamente da mesma forma que você usaria variáveis categóricas.

- Para testes de independência (qui-quadrado) ou comparação de proporções da coluna (testes z), os testes são executados em contagens, e a Contagem deve ser uma das estatísticas básicas exibidas na tabela.
- Para conjuntos de categorias múltiplas, os testes de comparação de proporções da coluna ou de médias de coluna (testes t) não serão executados se **Contar respostas duplicadas para conjuntos de categorias múltiplas** estiver selecionado na guia Opções. Consulte o tópico “Tabelas customizadas: Guia Opções” na página 16 para obter mais informações

Testes de independência com conjuntos de múltiplas respostas

Esse exemplo cria uma tabulação cruzada de uma variável categórica e um conjunto de múltiplas respostas e executa um teste qui-quadrado de independência na tabulação cruzada.

1. Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas).
2. Clique em **Reconfigurar** para limpar as configurações anteriores.
3. Arraste e solte *Fontes de notícias* (esse é o rótulo descritivo para o conjunto de múltiplas dicotomias \$mltnews) da lista de variáveis para a área Colunas da área de janela de tela.
4. Arraste e solte *Gênero* da lista de variáveis para a área Linhas da área de janela de tela.
5. Clique na guia **Estatística de teste**.
6. Selecione (marque) **Testes de independência (qui-quadrado)**.
7. Se ainda não estiver selecionado, selecione **Incluir variáveis de múltiplas respostas no teste**.
8. Clique em **OK** para executar o procedimento.

		News sources
Gender	Chi-square	10.266
	df	5
	Sig.	.068

Figura 73. Resultados de qui-quadrado

O nível de significância de 0,068 para o teste qui-quadrado indica que homens e mulheres provavelmente não diferem de forma significativa em suas escolhas de fontes de notícias (supondo que você usa um valor de significância de 0,05 ou inferior como seu critério para determinar a significância estatística).

Comparando médias de coluna com conjuntos de múltiplas respostas

Esse exemplo calcula médias de uma variável de escala em categorias definidas por um conjunto de múltiplas respostas e compara cada média de categoria com cada uma das outras médias de categoria para diferenças significativas.

1. Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas).
2. Clique em **Reconfigurar** para limpar as configurações anteriores.
3. Arraste e solte *Fontes de notícias* (esse é o rótulo descritivo para o conjunto de múltiplas dicotomias \$mltnews) da lista de variáveis para a área Colunas da área de janela de tela.
4. Arraste e solte *Idade do respondente* na área Linhas da área de janela de tela.
5. Clique na guia **Estatística de teste**.
6. Selecione (marque) **Comparar médias de coluna (testes t)**.
7. Se ainda não estiver selecionado, selecione **Incluir variáveis de múltiplas respostas no teste**.
8. Clique em **OK** para executar o procedimento.

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	Mean	Mean	Mean	Mean	Mean
Age of respondent	52	40	48	40	40

Comparisons of Column Means

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	(A)	(B)	(C)	(D)	(E)
Age of respondent	B C D E		B D E		

Results are based on two-sided tests assuming equal variances with significance level 0.05. For each significant pair, the key of the smaller category appears under the category with larger mean.

Figura 74. Resultados de teste de significância

- Cada categoria do conjunto de múltiplas respostas é identificada por uma letra (A, B, C, D, E), e para cada categoria para a qual a média de outra categoria é inferior e difere significativamente da média dessa categoria, a letra que representa a categoria com a média inferior é exibida.
- *Obter notícias de jornais* (A) tem a média de idade mais alta e todas as outras médias de categoria diferem significativamente dela.
- *Obter notícias da televisão* (C) tem a próxima média de idade mais alta, e todas as médias de categoria restantes (B, D e E) diferem significativamente dela. (C também difere significativamente de A, conforme indicado anteriormente.)
- As médias de idades para *Obter notícias de revistas* (B), *Obter notícias do rádio* (D) e *Obter notícias da Internet* (E) não diferem significativamente umas das outras.

Valores omissos

Muitos arquivos de dados contêm uma determinada quantidade de dados omissos. Uma ampla variedade de fatores pode resultar em dados omissos. Por exemplo, os respondentes da pesquisa de opinião podem não responder a todas as perguntas, algumas variáveis podem não ser aplicáveis a alguns casos, e erros de codificação podem resultar em alguns valores sendo lançados.

Há dois tipos de valores omissos no IBM SPSS Statistics:

- **Omisso de usuário.** Valores definidos como contendo dados omissos. Os rótulos de valor podem ser designados a esses valores para identificar por que os dados são omissos (como um código de 99 e um rótulo de valor de *Não aplicável* para gravidez em homens).
- **Omisso do sistema.** Se nenhum valor estiver presente para uma variável numérica, ela será designada ao valor omissos do sistema. Isso é indicado por um ponto na Visualização de dados do Editor de dados.

Há diversas instalações que podem ajudar a compensar os efeitos de dados omissos e até mesmo analisar padrões em dados omissos. No entanto, esse capítulo tem um objetivo muito mais simples: descrever como as Tabelas customizadas tratam dados omissos e como os dados omissos afetam o cálculo de estatísticas básicas.

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *missing_values.sav*. Consulte o tópico arquivos de dados para obter mais informações. Esse é um arquivo de dados muito simples, completamente artificial, com apenas uma variável e dez casos, projetado para ilustrar conceitos básicos sobre valores omissos.

Tabelas sem valores omissos

Por padrão, as categorias com usuário desconhecido não são exibidas em tabelas customizadas (e os valores omissos do sistema nunca são exibidos).

1. Nos menus, escolha:

Analisar > Tabelas > Tabelas Customizadas...

2. No construtor de tabela, arraste e solte *Variável com valores omissos* (a única variável no arquivo) da lista de variáveis para a área Linhas da área de janela de tela.
3. Clique com o botão direito na variável na área de janela de tela e selecione **Categorias e totais** no menu pop-up.
4. Clique em (marque) **Total** na caixa de diálogo Categorias e totais e, em seguida, clique em **Aplicar**.
5. Clique com o botão direito em *Variável com valores omissos* na visualização da tabela na área de janela de tela novamente e selecione **estatísticas básicas** no menu pop-up.
6. Na caixa de diálogo estatísticas básicas, selecione **% de N da coluna** na lista Estatísticas e clique na seta para incluí-la na lista Exibição.
7. Clique em **Aplicar à seleção**.

É possível notar uma pequena discrepância entre as categorias exibidas na visualização da tabela na área de janela de tela e as categorias exibidas na lista Categorias (abaixo da lista de variáveis, do lado esquerdo do construtor de tabela). A lista Categorias contém uma categoria rotulada *Valores omissos* que não está incluída na visualização da tabela, porque as categorias de valor omissos são excluídas por padrão. Como "valores" está no plural no rótulo, isso indica que a variável tem duas ou mais categorias com usuário desconhecido.

8. Clique em **OK** para criar a tabela.

		Count	Column N %
Variable with missing values	Low	2	28.6%
	Medium	3	42.9%
	High	2	28.6%
	Total	7	100.0%

Figura 75. Tabela sem valores omissos

Tudo nesta tabela está perfeitamente bem. Os valores de categoria são incluídos nos totais e as porcentagens refletem exatamente os valores obtidos usando a contagem total como a base de porcentagem (por exemplo, $3/7 = 0,429$ ou 42,9%). No entanto, a contagem total não é o número total de casos no arquivo de dados; é o número total de casos com valores **não omissos**, ou casos que não possuem valores omissos de usuário ou omissos do sistema para essa variável.

Incluindo valores omissos em tabelas

1. Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas).
2. Clique com o botão direito em *Variável com valores omissos* na visualização da tabela na área de janela de tela e selecione **Categorias e totais** no menu pop-up.
3. Clique em (marque) **Valores omissos** na caixa de diálogo Categorias e totais e, em seguida, clique em **Aplicar**.

Agora a visualização da tabela inclui uma categoria *Valores omissos*. Embora a visualização da tabela exiba apenas uma categoria para valores omissos, todas as categorias com usuário desconhecido serão exibidas na tabela.

4. Clique com o botão direito em *Variável com valores omissos* na visualização da tabela na área de janela de tela novamente e selecione **estatísticas básicas** no menu pop-up.
5. Na caixa de diálogo estatísticas básicas, clique em (marque) **estatísticas básicas customizada para totais e subtotais**.

6. Selecione **N válido** na lista estatísticas básicas customizada e clique na seta para incluí-lo na lista Exibição.
7. Faça o mesmo para **N Total**.
8. Clique em **Aplicar à seleção** e, em seguida, clique em **OK** no construtor de tabela para criar a tabela.

		Count	Column N %	Valid N	Total N
Variable with missing values	Low	2	22.2%		
	Medium	3	33.3%		
	High	2	22.2%		
	Don't know	1	11.1%		
	Not applicable	1	11.1%		
	Total	9	100.0%	7	10

Figura 76. Tabela com valores omissos

As duas categorias com usuário desconhecido definidas -- *Não sabe* e *Não aplicável* -- agora são exibidas na tabela, e a contagem total agora é 9 em vez de 7, refletindo a inclusão dos dois casos com valores omissos de usuário (um em cada categoria com usuário desconhecido). As porcentagens da coluna também são diferentes agora, porque elas são baseadas no número de valores não omissos e omissos de usuário. Apenas valores omissos do sistema não são incluídos no cálculo da porcentagem.

N válido mostra o número total de casos não omissos (7), e *N total* mostra o número total de casos, incluindo omissos de usuário e omissos do sistema. O número total de casos é 10, um a mais que a contagem de valores não omissos e omissos de usuário exibidos como o total na coluna *Contagem*. Isso ocorre porque há um caso com um valor omissos do sistema.

9. Abra o construtor de tabela (menu Analisar, Tabelas, Tabelas customizadas).
10. Clique com o botão direito em *Variável com valores omissos* na visualização da tabela na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
11. Selecione **% de N válido da coluna** na lista principal Estatísticas (não as sumarizações customizadas para totais e subtotaís) e clique na seta para incluí-la na lista Exibição.
12. Faça o mesmo para **% de N total da coluna**.
13. Também é possível incluí-los na lista de estatísticas básicas customizada para totais e subtotaís.
14. Clique em **Aplicar à seleção** e, em seguida, clique em **OK** para criar a tabela.

		Count	Column N %	Column Valid N %	Column Total N %	Valid N	Total N
Variable with missing values	Low	2	22.2%	28.6%	20.0%		
	Medium	3	33.3%	42.9%	30.0%		
	High	2	22.2%	28.6%	20.0%		
	Don't know	1	11.1%	.0%	10.0%		
	Not applicable	1	11.1%	.0%	10.0%		
	Total	9	100.0%	100.0%	100.0%	7	10

Figura 77. Tabela com valores omissos e porcentagens válidas e totais

- *% de N da coluna* é a porcentagem em cada categoria com base no número de valores não omissos e omissos de usuário (pois os valores omissos de usuário foram explicitamente incluídos na tabela).
- *% de N válido da coluna* é a porcentagem em cada categoria com base apenas nos casos válidos, não omissos. Esses valores são iguais às porcentagens da coluna na tabela original que não incluíam valores omissos de usuário.
- *% de N do total da coluna* é a porcentagem em cada categoria com base em todos os casos, incluindo omissos de usuário e omissos do sistema. Se você incluir as porcentagens de categorias individuais nessa categoria, verá que elas são incluídas somente até 90%, porque um caso do total de 10 casos (10%) possui o valor omissos do sistema. Embora esse caso seja incluído na base para os cálculos de porcentagem, nenhuma categoria é fornecida na tabela para casos com valores omissos do sistema.

Formatando e customizando tabelas

Formatando e customizando tabelas

As Tabelas customizadas permitem controlar várias propriedades de formatação de tabela como parte do processo de construção de tabela, incluindo:

- Formato de exibição e rótulos para estatísticas básicas
- Larguras mínima e máxima da coluna de dados
- Texto ou valor exibido em células vazias

Essas configurações persistem na interface do construtor de tabela (até que você as mude, reconfigure as configurações do construtor de tabela ou abra um arquivo de dados diferente), permitindo criar várias tabelas com as mesmas propriedades de formatação, sem editar manualmente as tabelas depois de criá-las. Também é possível salvar essas configurações de formatação, juntamente com todos os outros parâmetros da tabela, usando o botão Colar na interface do construtor de tabela para colar a sintaxe de comando em uma janela de sintaxe, que pode então ser salva como um arquivo.

Também é possível mudar muitas propriedades de formatação de tabelas após sua criação, usando todos os recursos de formatação disponíveis no Visualizador para tabelas dinâmicas. No entanto, esse capítulo focaliza o controle de propriedades de formatação de tabela antes da criação da tabela. Para obter informações adicionais sobre tabelas dinâmicas, use a guia Índice no sistema de Ajuda e digite tabelas dinâmicas como a palavra-chave.

Arquivo de dados de amostra

Os exemplos nesse capítulo usam o arquivo de dados *survey_sample.sav*. Consulte o tópico arquivos de dados para obter mais informações

Todos os exemplos fornecidos aqui exibem rótulos de variáveis em caixas de diálogo, ordenados em ordem alfabética. As propriedades de exibição da lista de variáveis são configuradas na guia Geral na caixa de diálogo Opções (menu Editar, Opções).

Formato de exibição de estatísticas básicas

As Tabelas customizadas tentam aplicar formatos padrão relativamente inteligentes a estatísticas básicas, mas provavelmente, algumas vezes você desejará substituir esses padrões.

1. Nos menus, escolha:

Analisar > Tabelas > Tabelas Customizadas...

2. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.
3. Arraste e solte *Confiança na televisão*, abaixo de *Categoria de idade* na área Linhas, empilhando as duas variáveis na dimensão da linha.
4. Clique com o botão direito em *Categoria de idade*, na visualização da tabela na área de janela de tela e selecione **Selecionar todas as variáveis de linha** no menu pop-up.
5. Clique com o botão direito em *Categoria de idade* novamente e selecione **Categorias e totais** no menu pop-up.
6. Na caixa de diálogo Categorias e totais, selecione (marque) **Total** e, em seguida, clique em **Aplicar**.
7. Clique com o botão direito em qualquer variável na visualização da tabela na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.
8. Selecione **% de N da coluna** na lista Estatísticas e clique na tecla de seta para incluí-la na lista Exibição.
9. Selecione (marque) **estatísticas básicas customizadas para totais e subtotais**.

10. Na lista Estatísticas para estatísticas básicas customizadas, selecione **% de N da coluna** e clique na seta para incluí-la na lista Exibição.
11. Faça o mesmo para **Média**.
12. Em seguida, clique em **Aplicar a todos**.

Os valores de item temporário na visualização da tabela refletem o formato padrão para cada estatística de sumarização.

- Para contagens, o formato de exibição padrão é **nnnn** -- valores de número inteiro sem casas decimais.
- Para porcentagens, o formato de exibição padrão é **nnnn.n%** -- números com uma única casa decimal e um sinal de porcentagem após o valor.
- Para a média, o formato de exibição padrão é *diferente* para as duas variáveis.

Para estatísticas básicas que não são alguma forma de contagem (incluindo N válido e N total) ou porcentagem, o formato de exibição padrão é o formato de exibição definido para a variável no Editor de dados. Se você examinar as variáveis na Visualização de variável no Editor de dados, verá que *Categoria de idade* (variável *agecat*) está definida como tendo duas casas decimais, enquanto *Confiança na televisão* (variável *contv*) está definida como tendo zero casas decimais.

Esse é um dos casos em que o formato padrão provavelmente não é o formato desejado pois, provavelmente, seria melhor se ambos os valores médios exibissem o mesmo número de decimais.

13. Clique com o botão direito em qualquer variável na visualização da tabela na área de janela de tela e selecione **estatísticas básicas** no menu pop-up.

Para a média, a célula Formato na lista Exibição indica que o formato é *Automático*, o que significa que o formato de exibição definido para a variável será usado, e a célula Casas Decimais está desativada. Para especificar o número de decimais, primeiro é necessário selecionar um formato diferente.

14. Na lista Exibição de estatísticas básicas customizada, clique na célula Formato para a média e selecione **nnnn** na lista suspensa de formatos.
15. Na célula Casas Decimais, insira um valor de 1.
16. Em seguida, clique em **Aplicar a todos** para aplicar esta configuração a ambas as variáveis.

Agora a visualização da tabela indica que os dois valores médios serão exibidos com uma casa decimal. (É possível prosseguir e criar essa tabela agora -- mas talvez você possa achar o valor "médio" para *Categoria de idade* um pouco difícil de interpretar, pois os códigos numéricos reais para essa variável variam de 1 a 6.)

Rótulos de exibição para estatísticas básicas

Além dos formatos de exibição para estatísticas básicas, também é possível controlar os rótulos descritivos para cada estatística de sumarização.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Clique em **Reconfigurar** para limpar todas as configurações anteriores no construtor de tabela.
3. No construtor de tabela, arraste e solte *Categoria de idade* da lista de variáveis para a área Linhas na área de janela de tela.
4. Arraste e solte *Como foi pago na semana passada* da lista de variáveis para a área Colunas na área de janela de tela.
5. Clique com o botão direito em *Categoria de idade*, na visualização da tabela da área de janela de tela e selecione **estatísticas básicas** no menu de contexto pop-up.
6. Selecione **% de N da coluna** na lista Estatísticas e clique na tecla de seta para incluí-la na lista Exibição.
7. Clique duas vezes em qualquer lugar na palavra *Coluna* na célula Rótulo na lista Exibição para editar o conteúdo da célula. Exclua a palavra *Coluna* do rótulo, mudando o rótulo para somente %.
8. Edite a célula Rótulo para *Contagem* da mesma maneira, mudando o rótulo para somente *N*.

Enquanto estamos aqui, vamos mudar o formato da estatística % de N da coluna para remover o sinal de porcentagem desnecessário (desde que o rótulo da coluna indique que a coluna contém porcentagens).

- Clique na célula Formato para % de N da coluna e selecione **nnnn.n** na lista suspensa de formatos.
- Em seguida, clique em **Aplicar à seleção**.

A visualização da tabela exibe o formato de exibição modificado e os rótulos modificados.

- Clique em **OK** para criar a tabela.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Figura 78. Tabela com rótulos de estatísticas básicas modificados

Largura da coluna

Você pode ter notado que a tabela no exemplo anterior é um tanto extensa. Uma solução para esse problema seria simplesmente trocar as variáveis de linha e de coluna. Outra solução é estreitar mais as colunas, já que elas parecem ser muito mais largas do que o necessário. (De fato, a razão pela qual encurtamos os rótulos de estatísticas básicas era para podermos estreitar mais as colunas.)

- Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
- Clique na guia **Opções**.
- No grupo Largura para colunas de dados, selecione **Customizado**.
- Para o Máximo, digite 36. (Certifique-se de que a configuração de Unidades seja **Pontos**.)
- Clique em **OK** para criar a tabela.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Figura 79. Tabela com larguras de colunas reduzidas

Agora a tabela é muito mais compacta.

Valor de exibição para células vazias

Por padrão, um 0 é exibido em células vazias (células que não contêm casos). Em vez disso, é possível não exibir nada nessas células (deixá-las em branco) ou especificar uma sequência de caracteres de texto para exibição em células vazias.

- Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
- Clique na guia **Opções**.
- No grupo Aparência da célula de dados, para Células vazias, selecione **Texto** e digite Nenhum.
- Clique em **OK** para criar a tabela.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	None	None	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	None	None	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	0.9	15	8.3

Figura 80. Tabela com "Nenhum" exibido em células vazias

Agora as quatro células vazias na tabela exibem o texto *Nenhum* em vez de um valor 0.

Valor de exibição para estatísticas omissas

Se uma estatística não puder ser calculada, o valor de exibição padrão será um ponto (.), que é o símbolo usado para indicar o valor omissos do sistema. Isso é diferente de uma célula "vazia" e, portanto, o valor de exibição para estatísticas omissas é controlado separadamente do valor de exibição para células que não contêm casos.

1. Abra o construtor de tabela (Menu Analisar, Tabelas, tabelas Customizadas).
2. Arraste e solte *Horas por dia assistindo à TV* da lista de variáveis para a parte superior da área Colunas na tela, acima de *Como foi pago na semana passada*.
Como *Horas por dia assistindo à TV* é uma variável de escala, ela se torna automaticamente a variável de origem estatística e a estatística de sumarização muda para a média.
3. Clique com o botão direito em *Horas por dia assistindo à TV* na visualização da tabela na área de janela de tela e selecione **estatísticas básicas** no menu de contexto pop-up.
4. Selecione **N válido** na lista Estatísticas e clique na tecla de seta para incluí-lo na lista Exibição.
5. Clique em **Aplicar à seleção**.
6. Clique na guia **Opções**.
7. No campo de texto para Estatísticas que não podem ser calculadas, digite NA.
8. Clique em **OK** para criar a tabela.

		Hours per day watching TV											
		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N
Age category	Less than 25	3	71	NA	None	3	10	2	3	2	6	2	8
	25 to 34	3	134	5	2	2	30	2	29	2	52	2	22
	35 to 44	3	136	2	5	3	30	2	34	2	47	3	46
	45 to 54	2	90	2	4	2	22	2	36	2	45	2	34
	55 to 64	3	40	NA	None	3	7	2	15	2	23	3	15
	65 or older	3	18	2	2	1	1	NA	0	1	2	3	11

Figura 81. Tabela com "NA" exibido para estatísticas omissas

O texto NA é exibido para a média em três células na tabela. Em cada caso, o valor correspondente *N válido* explica o motivo: Não há casos com os quais calcular a média.

No entanto, você pode notar o que parece ser uma pequena discrepância — um desses três valores *N válido* é exibido como um 0, em vez do rótulo *Nenhum*, que deve ser exibido nas células sem nenhum caso. Isso ocorre porque, embora não existam casos válidos a serem usados para calcular a média, a categoria não está realmente vazia. Se você voltar à tabela original com apenas as duas variáveis categóricas, verá que existem, de fato, três casos nessa categoria de tabulação cruzada. No entanto, não existem casos válidos, porque todos os três possuem valores omissos para a variável de escala *Horas por dia assistindo à TV*.

Arquivos de Amostra

Os arquivos de amostra instalados com o produto podem ser localizados no subdiretório *Amostras* do diretório de instalação. Há uma pasta separada dentro do subdiretório Amostras para cada um dos seguintes idiomas: inglês, francês, alemão, italiano, japonês, coreano, polonês, russo, chinês simplificado, espanhol e chinês tradicional.

Nem todos os arquivos de amostra estão disponíveis em todos os idiomas. Se um arquivo de amostra não estiver disponível em um idioma, essa pasta de idiomas conterá uma versão em inglês do arquivo de amostra.

PDV

A seguir, encontram-se breves descrições dos arquivos de amostra usadas em vários exemplos em toda a documentação.

- **accidents.sav.** Este é um arquivo de dados hipotéticos que diz respeito a uma empresa de seguros que está estudando fatores de risco de idade e sexo para acidentes de automóveis em uma determinada região. Cada caso corresponde a uma classificação cruzada de categoria de idade e sexo.
- **adl.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços para determinar os benefícios de um tipo proposto de terapia para pacientes que sofreram acidente vascular cerebral. Os médicos designaram aleatoriamente pacientes do sexo feminino que sofreram acidente vascular cerebral a um dos dois grupos. O primeiro recebeu a terapia física padrão e o segundo recebeu uma terapia emocional adicional. Três meses seguindo os tratamentos, as habilidades de cada paciente em realizar atividades comuns da vida diária foram pontuadas como variáveis ordinais.
- **advert.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços do varejista em examinar o relacionamento entre o dinheiro gasto em publicidade e as vendas resultantes. Para essa finalidade, eles coletaram dados de vendas passadas e os custos associados de publicidade.
- **aflatoxin.sav.** Este é um arquivo de dados hipotéticos que diz respeito ao teste de safras de milho para aflatoxina, um veneno cuja concentração varia muito entre e dentro das produções da safra. Um processador de grãos recebeu 16 amostras de cada uma das 8 produções da safra e mediu os níveis de aflatoxina em partes por bilhão (PPB).
- **anorectic.sav.** Enquanto trabalhando em direção a uma sintomatologia padronizada do comportamento anorético/bulímico, os pesquisadores ¹ fizeram um estudo de 55 adolescentes com transtornos alimentares conhecidos. Cada paciente foi visto quatro vezes durante quatro anos, para um total de 220 observações. Em cada observação, os pacientes foram pontuados para cada um dos 16 sintomas. As pontuações dos sintomas estão faltando para o paciente 71 no tempo 2, paciente 76 no tempo 2 e paciente 47 no tempo 3, deixando 217 observações válidas.
- **bankloan.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de um banco em reduzir a taxa de padrões de empréstimo. O arquivo contém informações financeiras e demográficas sobre 850 clientes passados e potenciais. Os primeiros 700 casos são clientes que anteriormente receberam empréstimos. Os últimos 150 casos são clientes potenciais que o banco precisa classificar os riscos de crédito como bons ou ruins.
- **bankloan_binning.sav.** Este é um arquivo de dados hipotéticos que contém informações financeiras e demográficas sobre 5.000 clientes antigos.
- **behavior.sav.** Em um exemplo clássico ², foi solicitado que 52 estudantes classificassem as combinações de 15 situações e 15 comportamentos em uma escala de 10 pontos variando de 0="extremamente adequado" a 9="extremamente inadequado." Na média de indivíduos, os valores são assumidos como dissimilaridades.

1. Van der Ham, T., J. J. Meulman, D. C. Van Strien e H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363-368.

2. Price, R. H., e D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579-586.

- **behavior_ini.sav.** Este arquivo de dados contém uma configuração inicial para uma solução bidimensional para *behavior.sav*.
- **brakes.sav.** Este é um arquivo de dados hipotéticos que diz respeito ao controle de qualidade em uma fábrica que produz freios de disco para automóveis de alto desempenho. O arquivo de dados contém medidas de diâmetro de 16 discos de cada uma das 8 máquinas de produção. O diâmetro de destino para os freios é de 322 milímetros.
- **breakfast.sav.** Em um estudo clássico ³, foi solicitado que 21 estudantes de MBA da Wharton School e seus cônjuges classificassem 15 itens de café da manhã em ordem de preferência com 1="mais preferencial" a 15="menos preferencial". Suas preferências foram registradas em seis cenários diferentes, de "Preferência geral" a "Petiscos, apenas com bebidas".
- **breakfast-overall.sav.** Este arquivo de dados contém as preferências dos itens de café da manhã para o primeiro cenário, "Preferência geral", apenas.
- **broadband_1.sav.** Este é um arquivo de dados hipotéticos que contém o número de assinantes, por região, de um serviço nacional de banda larga. O arquivo de dados contém os números de assinantes mensais para 85 regiões durante um período de quatro anos.
- **broadband_2.sav.** Este arquivos de dados é idêntico para *broadband_1.sav*, mas contém dados para três meses adicionais.
- **car_insurance_claims.sav.** Um conjunto de dados apresentado e analisado em outro lugar ⁴ diz respeito a reclamações por danos nos carros. A quantia média de reclamações pode ser modelada como tendo uma distribuição gama, usando uma função de ligação inversa para relacionar a média da variável dependente a uma combinação linear da idade do beneficiário do seguro, do tipo de veículo e da idade do veículo. O número de reclamações arquivadas pode ser usado como um peso de ajuste de escala.
- **car_sales.sav.** Este arquivo de dados contém estimativas de vendas hipotéticas, preços de lista e especificações físicas para várias marcas e modelos de veículos. Os preços de listas e as especificações físicas foram obtidos alternadamente de *edmunds.com* e de sites do fabricante.
- **car_sales_uprepared.sav.** Esta é uma versão modificada de *car_sales.sav* que não inclui nenhuma versão transformada dos campos.
- **carpet.sav.** Em um exemplo popular ⁵, uma empresa interessada em fazer propaganda de um novo limpador de carpete quer examinar a influência de cinco fatores de preferência do consumidor—design de pacote, nome de marca, preço, um selo *Good Housekeeping* e garantia de retorno financeiro. Há três níveis de fatores para design de pacote, cada um diferindo no local da escova do aplicador; três nomes de marcas (*K2R*, *Glory* e *Bissell*); três níveis de preço e dois níveis (não ou sim) para cada um dos dois últimos fatores. Dez consumidores classificam 22 perfis definidos por esses fatores. A variável *Preferência* contém a classificação média para cada perfil. Classificações baixas correspondem à preferência alta. Esta variável reflete uma medida geral de preferência para cada perfil.
- **carpet_prefs.sav.** Este arquivo de dados é baseado no mesmo exemplo descrito para *carpet.sav*, mas contém as classificações reais coletadas de cada um dos 10 consumidores. Foi solicitado que os consumidores classificassem os 22 perfis de produto do mais preferencial ao menos preferencial. As variáveis *PREF1* a *PREF22* contém os identificadores dos perfis associados, conforme definido em *carpet_plan.sav*.
- **catalog.sav.** Este arquivo de dados contém dados de vendas mensais hipotéticos para três produtos vendidos por uma empresa de catálogo. Dados para cinco variáveis preditoras possíveis também estão incluídos.
- **catalog_seasfac.sav.** Este arquivo de dados é o mesmo que *catalog.sav*, exceto pela adição de um conjunto de fatores sazonais calculados a partir do procedimento de Decomposição Sazonal juntamente com as variáveis de data que acompanham.

3. Green, P. E., e V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

4. McCullagh, P., e J. A. Nelder. 1989. *Generalized Linear Models*, 2ª ed. Londres: Chapman & Hall.

5. Green, P. E., e Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

- **cellular.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de uma empresa de telefonia celular para reduzir a perda de clientes. Os escores de propensão à perda de clientes são aplicados às contas, variando de 0 a 100. A pontuação de contas de 50 ou acima pode estar buscando a mudança dos provedores.
- **ceramics.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de um fabricante em determinar se uma nova liga premium tem uma resistência maior ao calor do que uma liga padrão. Cada caso representa um teste separado de uma das ligas; o calor no qual o rolamento falhou é registrado.
- **cereal.sav.** Este é um arquivo de dados hipotéticos que diz respeito a uma pesquisa com 880 pessoas sobre suas preferências no café da manhã, observando também sua idade, sexo, estado civil e se elas têm ou não um estilo de vida ativo (com base em se elas se exercitam pelo menos duas vezes por semana). Cada caso representa um entrevistado separado.
- **clothing_defects.sav.** Este é um arquivo de dados hipotéticos que diz respeito ao processo de controle de qualidade em uma fábrica de vestuário. De cada lote produzido na fábrica, os inspetores pegam uma amostra de roupas e contam o número de peças que não são aceitáveis.
- **coffee.sav.** Este arquivo de dados pertence às imagens percebidas de seis marcas de café gelado ⁶. Para cada um dos 23 atributos de imagens de café gelado, as pessoas selecionaram todas as marcas que foram descritas pelo atributo. As seis marcas são indicadas como AA, BB, CC, DD, EE e FF para preservar a confidencialidade.
- **contacts.sav.** Este é um arquivo de dados hipotéticos que diz respeito às listas de contato para um grupo de representantes de vendas de computadores corporativos. Cada contato é categorizado pelo departamento da empresa em que trabalham e suas classificações da empresa. Também registrada é a quantia da última venda realizada, o tempo desde a última venda e o tamanho da empresa de contato.
- **creditpromo.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de uma loja de departamentos para avaliar a eficácia de uma promoção recente de cartões de crédito. Para essa finalidade, 500 titulares de cartões foram selecionados aleatoriamente. Metade recebeu um anúncio promovendo uma taxa de juros reduzida em compras feitas nos próximos três meses. Metade recebeu um anúncio sazonal padrão.
- **customer_dbase.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de uma empresa em usar as informações em seu data warehouse para fazer ofertas especiais para clientes que têm maior probabilidade de responder. Um subconjunto da base de clientes foi selecionado aleatoriamente e recebeu as ofertas especiais e suas respostas foram registradas.
- **customer_information.sav.** Um arquivo de dados hipotéticos que contém informações de correspondência do cliente, como nome e endereço.
- **customer_subset.sav.** Um subconjunto de 80 casos de *customer_dbase.sav*.
- **debate.sav.** Este é um arquivo de dados hipotéticos que diz respeito às respostas pairwise a uma pesquisa de participantes de um debate político antes e depois do debate. Cada caso corresponde a um entrevistado separado.
- **debate_aggregate.sav.** Este é um arquivo de dados hipotéticos que agrega as respostas em *debate.sav*. Cada caso corresponde a uma classificação cruzada de preferência antes e depois do debate.
- **demo.sav.** Este é um arquivo de dados hipotéticos que diz respeito a um banco de dados de clientes adquiridos, com o objetivo de enviar por e-mail as ofertas mensais. É registrado se o cliente respondeu ou não à oferta, juntamente com várias informações demográficas.
- **demo_cs_1.sav.** Este é um arquivo de dados hipotéticos que diz respeito à primeira etapa dos esforços de uma empresa em compilar um banco de dados de informações de pesquisas. Cada caso corresponde a uma cidade diferente e a região, província, distrito e identificação da cidade são registrados.
- **demo_cs_2.sav.** Este é um arquivo de dados hipotéticos que diz respeito à segunda etapa dos esforços de uma empresa em compilar um banco de dados de informações de pesquisas. Cada caso corresponde a uma unidade doméstica diferente das cidades selecionadas na primeira etapa e a região, província,

6. Kennedy, R., C. Riquier, e B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56-70.

distrito, cidade, subdivisão e identificação da unidade são registrados. As informações de amostragem dos dois primeiros estágios do design também são incluídas.

- **demo_cs.sav.** Este é um arquivo de dados hipotéticos que contém informações de pesquisas coletadas usando um plano de amostragem complexo. Cada caso corresponde a uma unidade doméstica diferente e várias informações demográficas e de amostragem são registradas.
- **diabetes_costs.sav.** Este é um arquivo de dados hipotéticos que contém informações que são mantidas por uma empresa de seguros sobre beneficiários do seguro que têm diabetes. Cada caso corresponde a um beneficiário de seguro diferente.
- **dietstudy.sav.** Este arquivo de dados hipotéticos contém os resultados de um estudo da "dieta Stillman"⁷. Cada caso corresponde a um assunto separado e registra os pesos em libras antes e depois da dieta e os níveis de triglicérides em mg/100 ml.
- **dmdata.sav.** Este é um arquivo de dados hipotéticos que contém informações demográficas e de compras para uma empresa de marketing direto. *dmdata2.sav* contém informações para um subconjunto de contatos que recebeu um teste de envio e *dmdata3.sav* contém informações sobre os contatos restantes que não receberam o teste de envio.
- **dvdplayer.sav.** Este é um arquivo de dados hipotéticos que diz respeito ao desenvolvimento de um novo DVD player. Usando um protótipo, a equipe de marketing coletou dados do grupo em foco. Cada caso corresponde a um usuário pesquisado separado e registra algumas informações demográficas sobre ele e suas respostas às perguntas sobre o protótipo.
- **german_credit.sav.** Este arquivo de dados é obtido do conjunto de dados "Crédito alemão" no Repositório de Bancos de Dados de Aprendizado de Máquina⁸ na Universidade da Califórnia, Irvine.
- **grocery_1month.sav.** Este arquivo de dados hipotéticos é o arquivo de dados *grocery_coupons.sav* com as compras semanais "agregadas" para que cada caso corresponda a um cliente separado. Algumas das variáveis que foram alteradas semanalmente desaparecem como resultado e o montante gasto registrado é, agora, a soma dos montantes gastos durante as quatro semanas do estudo.
- **grocery_coupons.sav.** Este é um arquivo de dados hipotéticos que contém dados da pesquisa coletados por uma cadeia de supermercados interessada nos hábitos de compra de seus clientes. Cada cliente é seguido durante quatro semanas e cada caso corresponde às informações de registros e semanais do cliente separado sobre onde e como o cliente compra, incluindo quanto foi gasto em supermercados durante essa semana.
- **guttman.sav.** Bell⁹ apresentou uma tabela para ilustrar os grupos sociais possíveis. Guttman¹⁰ usou uma parte dessa tabela, na qual cinco variáveis descrevendo coisas como interação social, sentimentos em pertencer a um grupo, proximidade física dos membros e formalidade do relacionamento foram cruzadas com sete grupos sociais teóricos, incluindo multidões (por exemplo, pessoas em um jogo de futebol), públicos (por exemplo, pessoas em um teatro ou leitura em sala de aula), público (por exemplo, públicos de jornais ou televisão), aglomerações (como uma multidão, mas com interação muito mais intensa), grupos primários (familiares), grupos secundários (voluntários) e a comunidade moderna (confederação informal resultante da grande proximidade física e da necessidade de serviços especializados).
- **health_funding.sav.** Este é um arquivo de dados hipotéticos que contém dados sobre financiamento de saúde (quantia por 100 da população), taxas de doenças (taxa por 10.000 da população) e visitas aos provedores de assistência médica (taxa por 10.000 da população). Cada caso representa uma cidade diferente.
- **hivassay.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de um laboratório farmacêutico em desenvolver uma análise rápida para detectar a infecção por HIV. Os resultados da

7. Rickman, R., N. Mitchell, J. Dingman, e J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228:, 54-58.

8. Blake, C. L., e C. J. Merz. 1998. "UCI Repository of machine learning databases". Disponível em <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

9. Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. Nova York: Harper & Row.

10. Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469-506.

análise são oito tonalidades profundas de vermelho, com tonalidades mais profundas indicando maior probabilidade de infecção. Um teste de laboratório foi realizado em 2.000 amostras de sangue, metade das quais estava infectada com HIV e metade estava limpa.

- **hourlywagedata.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos salários por hora de enfermeiros de posições de escritório e hospitalar e com níveis variados de experiência.
- **insurance_claims.sav.** Este é um arquivo de dados hipotéticos que diz respeito a uma empresa de seguros que deseja construir um modelo para sinalizar reclamações suspeitas e potencialmente fraudulentas. Cada caso representa uma reclamação separada.
- **insure.sav.** Este é um arquivo de dados hipotéticos que diz respeito a uma empresa de seguros que está estudando os fatores de risco que indicam se um cliente terá que fazer uma reclamação sobre um contrato de seguro de vida de termo de 10 anos. Cada caso no arquivo de dados representa um par de contratos, um dos quais registrou uma reclamação e o outro não registrou, com correspondência de idade e sexo.
- **judges.sav.** Este é um arquivo de dados hipotéticos que diz respeito às pontuações fornecidas por juízes treinados (mais um entusiasta) a 300 apresentações de ginástica. Cada linha representa uma apresentação separada; os juízes visualizaram as mesmas apresentações.
- **kinship_dat.sav.** Rosenberg e Kim ¹¹ definiram analisar 15 termos de parentesco (tia, irmão, primo, filha, pai, neta, avô, avó, neto, mãe, sobrinho, sobrinha, irmã, filho, tio). Eles pediram a quatro grupos de estudantes universitários (dois femininos, dois masculinos) que classificassem esses termos com base em semelhanças. Foi solicitado que dois grupos (um feminino, um masculino) classificassem duas vezes, com a segunda classificação baseada em um critério diferente da primeira. Assim, um total de seis “fontes” foi obtido. Cada fonte corresponde a uma matriz de proximidade de 15 x 15, cujas células são iguais ao número de pessoas em uma fonte menos o número de vezes que os objetos foram particionados juntos nessa fonte.
- **kinship_ini.sav.** Este arquivo de dados contém uma configuração inicial para uma solução tridimensional para *kinship_dat.sav*.
- **kinship_var.sav.** Este arquivo de dados contém variáveis independentes *sexo*, *ger(ação)* e *grau* (de separação) que podem ser usadas para interpretar as dimensões de uma solução para *kinship_dat.sav*. Especificamente, elas podem ser usadas para restringir o espaço da solução para uma combinação linear dessas variáveis.
- **marketvalues.sav.** Este arquivo de dados diz respeito a vendas de moradias no desenvolvimento de um novo alojamento em Algonquin, Ill., durante os anos de 1999–2000. Estas vendas são uma questão de registro público.
- **nhis2000_subset.sav.** A National Health Interview Survey (NHIS) é uma pesquisa grande, baseada na população civil dos EUA. As entrevistas são realizadas através de comunicação direta em uma amostra nacionalmente representativa das famílias. Informações demográficas e observações sobre comportamentos de funcionamento e status são obtidas para membros de cada família. Esse arquivo de dados contém um subconjunto de informações da pesquisa de 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Documentação e arquivo de dados de uso público. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/. Acessado 2003.
- **ozone.sav.** Os dados incluem 330 observações sobre seis variáveis meteorológicas para prever a concentração de ozônio das variáveis restantes. Pesquisadores anteriores ^{12, 13}, entre outras coisas, encontraram falta de linearidade entre essas variáveis, o que prejudica as abordagens de regressão padrão.
- **pain_medication.sav.** Este arquivo de dados hipotéticos contém os resultados de um teste clínico para medicação anti-inflamatória para tratar a dor artrítica crônica. De interesse específico é o tempo que leva para o medicamento fazer efeito e como ele se compara a uma medicação existente.

11. Rosenberg, S., e M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489-502.

12. Breiman, L., e J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.

13. Hastie, T., e R. Tibshirani. 1990. *Generalized additive models*. Londres: Chapman e Hall.

- **patient_los.sav.** Este arquivo de dados hipotéticos contém os registros do tratamento de pacientes que foram internados no hospital por suspeita de enfarte do miocárdio (MI ou "ataque cardíaco"). Cada caso corresponde a um paciente separado e registra muitas variáveis relacionadas à sua permanência no hospital.
- **patlos_sample.sav.** Este arquivo de dados hipotéticos contém os registros de tratamento de uma amostra de pacientes que recebeu trombolíticos durante o tratamento para enfarte do miocárdio (MI ou "ataque cardíaco"). Cada caso corresponde a um paciente separado e registra muitas variáveis relacionadas à sua permanência no hospital.
- **poll_cs.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de especialistas em pesquisa em determinar o nível de suporte público para uma nota antes da legislatura. Os casos correspondem aos eleitores registrados. Cada caso registra o estado, município e bairro em que o eleitor vive.
- **poll_cs_sample.sav.** Este arquivo de dados hipotéticos contém uma amostra dos eleitores listados em *poll_cs.sav*. A amostra foi obtida de acordo com o design especificado no arquivo de plano *poll_csplan* e este arquivo de dados registra as probabilidades de inclusão e as ponderações da amostra. Observe, no entanto, que, como o plano de amostragem usa um método de probabilidade-proporcional-ao-tamanho (PPS), há também um arquivo que contém as probabilidades de seleção de conjunto (*poll_jointprob.sav*). As variáveis adicionais correspondentes aos dados demográficos do eleitor e sua opinião sobre a nota proposta foram coletadas e incluíram o arquivo de dados depois que a amostra foi obtida.
- **property_assess.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de um assessor de estado em manter as avaliações do valor de propriedade atualizadas em relação aos recursos limitados. Os casos correspondem às propriedades vendidas no estado no ano passado. Cada caso no arquivo de dados registra o município no qual a propriedade está, o assessor que visitou pela última vez a propriedade, o tempo desde essa avaliação, a valorização feita naquele momento e o valor de venda da propriedade.
- **property_assess_cs.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de um assessor de estado em manter as avaliações de valor da propriedade atualizadas em relação aos recursos limitados. Os casos correspondem às propriedades no estado. Cada caso no arquivo de dados registra o estado, município e bairro em que a propriedade está, o tempo desde a última avaliação e a valorização feita naquele momento.
- **property_assess_cs_sample.sav.** Este arquivo de dados hipotéticos contém uma amostra das propriedades listadas em *property_assess_cs.sav*. A amostra foi obtida de acordo com o design especificado no arquivo de plano *property_assess_csplan* e esse arquivo de dados registra as probabilidades de inclusão e as ponderações da amostra. A variável adicional *Valor atual* foi coletada e incluída no arquivo de dados depois que a amostra foi obtida.
- **recidivism.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de uma agência de aplicação da lei do governo em entender as taxas de reincidência em sua área de jurisdição. Cada caso corresponde a um ofensor anterior e registra suas informações demográficas, alguns detalhes de seu primeiro crime e o tempo até sua segunda prisão, se tiver ocorrido dentro de dois anos após a primeira prisão.
- **recidivism_cs_sample.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de uma agência de aplicação da lei do governo em entender as taxas de reincidência em sua área de jurisdição. Cada caso corresponde a um ofensor anterior, solto de sua primeira prisão durante o mês de junho de 2003 e registra suas informações demográficas, alguns detalhes de seu primeiro crime e os dados de sua segunda prisão, se tiver ocorrido no final de junho de 2006. Os ofensores foram selecionados de departamentos de amostragem de acordo com o plano de amostragem especificado em *recidivism_cs.csplan*; como ele faz uso de um método de probabilidade-proporcional-ao-tamanho (PPS), há também um arquivo que contém as probabilidades de seleção de conjunto (*recidivism_cs_jointprob.sav*).
- **rfm_transactions.sav.** Um arquivo de dados hipotéticos que contém dados da transação de compra, incluindo a data da compra, o(s) item(ns) comprado(s) e o valor monetário de cada transação.
- **salesperformance.sav.** Este é um arquivo de dados hipotéticos que diz respeito à avaliação de dois novos cursos de treinamento de vendas. Sessenta funcionários, divididos em três grupos, receberão

treinamento padrão. Além disso, o grupo 2 obtém treinamento técnico; o grupo 3, um tutorial prático. Cada funcionário foi testado no final do curso de treinamento e sua pontuação foi registrada. Cada caso no arquivo de dados representa um estagiário separado e registra o grupo ao qual ele foi designado e a pontuação que recebeu no exame.

- **satisf.sav.** Este é um arquivo de dados hipotéticos que diz respeito a uma pesquisa de satisfação realizada por uma empresa de varejo em 4 locais de loja. 582 clientes foram pesquisados e cada caso representa as respostas de um único cliente.
- **screws.sav.** Este arquivo de dados contém informações sobre as características de parafusos, cavilhas, porcas e tachas ¹⁴.
- **shampoo_ph.sav.** Este é um arquivo de dados hipotéticos que diz respeito ao controle de qualidade em uma fábrica de produtos para cabelos. Em intervalos de tempo regulares, seis lotes de saída separados são medidos e seu pH é registrado. O intervalo de destino é 4.5–5.5.
- **ships.sav.** Um conjunto de dados apresentado e analisado em outro lugar ¹⁵ que diz respeito aos danos aos navios de carga causados pelas ondas. As contagens de incidentes podem ser modeladas como ocorrendo em uma taxa de Poisson dado o tipo do navio, o período de construção e o período de serviço. Os meses agregados de serviço para cada célula da tabela formada pela classificação cruzada de fatores fornecem os valores para a exposição ao risco.
- **site.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de uma empresa em escolher novos sites para sua empresa em expansão. Eles contrataram dois consultores para avaliar separadamente os sites, quem, além de um relatório estendido, resumiu cada site como um possível cliente "bom", "justo" ou "fraco".
- **smokers.sav.** Este arquivo de dados é resumido do 1998 National Household Survey of Drug Abuse e é uma amostra da probabilidade dos lares americanos. (<http://dx.doi.org/10.3886/ICPSR02934>) Assim, a primeira etapa em uma análise deste arquivo de dados deve ser ponderar os dados para refletir as tendências da população.
- **stocks.sav** Este arquivo de dados hipotéticos contém preços e volume de estoque para um ano.
- **stroke_clean.sav.** Este arquivo de dados hipotéticos contém o estado de um banco de dados médico após ter sido limpo seguindo os procedimentos em Statistics Base Edition.
- **stroke_invalid.sav.** Este arquivo de dados hipotéticos contém o estado inicial de um banco de dados médico e contém vários erros de entrada de dados.
- **stroke_survival.** Este arquivo de dados hipotéticos diz respeito aos tempos de sobrevivência para pacientes que estão saindo de um programa de reabilitação pós-derrame isquêmico diante de vários desafios. Pós-derrame, a ocorrência de enfarte do miocárdio, derrame isquêmico ou derrame hemorrágico é observado e o momento do evento registrado. A amostra é truncada à esquerda porque inclui apenas pacientes que sobreviveram até o fim do programa de reabilitação administrado pós-derrame.
- **stroke_valid.sav.** Este arquivo de dados hipotéticos contém o estado de um banco de dados médico depois que os valores foram verificados usando o procedimento Validar Dados. Ele contém também casos potencialmente anormais.
- **survey_sample.sav.** Este arquivo de dados contém dados da pesquisa, incluindo dados demográficos e várias medidas de atitude. Ele é baseado em um subconjunto de variáveis do 1998 NORC General Social Survey, embora alguns valores de dados tenham sido modificados e variáveis fictícias adicionais tenham sido incluídas para fins de demonstração.
- **tcm_kpi.sav.** Este é um arquivo de dados hipotéticos que contém valores dos principais indicadores de desempenho semanais para um negócio. Ele também contém dados semanais para várias métricas controláveis durante o mesmo período de tempo.
- **tcm_kpi_upd.sav.** Este arquivo de dados é idêntico a *tcm_kpi.sav*, mas contém dados para quatro semanas extras.

14. Hartigan, J. A. 1975. *Clustering algorithms*. Nova York: John Wiley and Sons.

15. McCullagh, P., e J. A. Nelder. 1989. *Generalized Linear Models*, 2ª ed. Londres: Chapman & Hall.

- **telco.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de uma empresa de telecomunicações para reduzir a perda de clientes na base de clientes. Cada caso corresponde a um cliente separado e registra várias informações demográficas e de uso de serviço.
- **telco_extra.sav.** Este arquivo de dados é semelhante ao arquivo de dados *telco.sav*, mas o "aforamento" e as variáveis de gastos do cliente transformadas por log foram removidos e substituídos por variáveis de gasto do cliente transformadas por log padronizadas.
- **telco_missing.sav.** Este arquivo de dados é um subconjunto do arquivo de dados *telco.sav*, mas alguns dos valores de dados demográficos foram substituídos por valores omissos.
- **testmarket.sav.** Este arquivo de dados hipotéticos diz respeito aos planos de uma cadeia de fast food para incluir um novo item no menu. Há três campanhas possíveis para promover o novo produto, portanto, o novo item é introduzido em locais em vários mercados selecionados aleatoriamente. Uma promoção diferente é usada em cada local e as vendas semanais do novo item são registradas para as primeiras quatro semanas. Cada caso corresponde a um local semanal separado.
- **testmarket_1month.sav.** Este arquivo de dados hipotéticos é o arquivo de dados *testmarket.sav* com as vendas semanais "agregadas" para que cada caso corresponda a um local separado. Algumas das variáveis que foram alteradas semanalmente desaparecem como resultado e as vendas registradas são, agora, a soma das vendas durante as quatro semanas do estudo.
- **tree_car.sav.** Este é um arquivo de dados hipotéticos que contém dados demográficos e de preço de compra do veículo.
- **tree_credit.sav.** Este é um arquivo de dados hipotéticos que contém dados demográficos e de histórico de empréstimo bancário.
- **tree_missing_data.sav** Este é um arquivo de dados hipotéticos que contém dados demográficos e de histórico de empréstimo bancário com um grande número de valores omissos.
- **tree_score_car.sav.** Este é um arquivo de dados hipotéticos que contém dados demográficos e de preço de compra do veículo.
- **tree_textdata.sav.** Um arquivo de dados simples com apenas duas variáveis destinadas principalmente a mostrar o estado padrão de variáveis antes da designação do nível de medição e dos rótulos de valor.
- **tv-survey.sav.** Este é um arquivo de dados hipotéticos que diz respeito a uma pesquisa realizada por um estúdio de televisão que está considerando se deve estender a execução de um programa bem-sucedido. Foi perguntado a 906 entrevistados se eles assistiriam ao programa sob várias condições. Cada linha representa um entrevistado separado; cada coluna é uma condição separada.
- **ulcer_recurrence.sav.** Este arquivo contém informações parciais de um estudo projetado para comparar a eficácia de duas terapias para evitar a recorrência de úlceras. Ele fornece um bom exemplo de dados com intervalo censurado e foi apresentado e analisado em outro lugar ¹⁶.
- **ulcer_recurrence_recoded.sav.** Este arquivo reorganiza as informações em *ulcer_recurrence.sav* para permitir que você modele a probabilidade de eventos para cada intervalo do estudo em vez de simplesmente a probabilidade de eventos no final do estudo. Ele foi apresentado e analisado em outro lugar ¹⁷.
- **verd1985.sav.** Este arquivo de dados diz respeito a uma pesquisa ¹⁸. As respostas de 15 pessoas a 8 variáveis foram registradas. As variáveis de interesse são divididas em três conjuntos. O conjunto 1 inclui *idade* e *estado civil*, o conjunto 2 inclui *animal de estimação* e *notícias*, e o conjunto 3 inclui *música* e *em tempo real*. *Animal de estimação* é escalado como nominal múltiplo e *idade* é escalado como ordinal; todas as outras variáveis são escaladas como nominal único.

16. Collett, D. 2003. *Modelling survival data in medical research*, 2ª ed. Boca Raton: Chapman & Hall/CRC.

17. Collett, D. 2003. *Modelling survival data in medical research*, 2ª ed. Boca Raton: Chapman & Hall/CRC.

18. Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (em alemão)*. Leiden: Department of Data Theory, Universidade de Leiden.

- **virus.sav.** Este é um arquivo de dados hipotéticos que diz respeito aos esforços de um provedor de serviços de Internet (ISP) em determinar os efeitos de um vírus em suas redes. Eles têm controlado a porcentagem (aproximada) de tráfego de e-mail infectado em suas redes ao longo do tempo, do momento da descoberta até a ameaça ser contida.
- **wheeze_steubenville.sav.** Este é um subconjunto de um estudo longitudinal dos efeitos da poluição do ar na saúde de crianças ¹⁹. Os dados contêm medidas binárias repetidas de status de respiração difícil para crianças de Steubenville, Ohio, nas idades de 7, 8, 9 e 10 anos, juntamente com um registro fixo de se a mãe era ou não fumante durante o primeiro ano do estudo.
- **workprog.sav.** Este é um arquivo de dados hipotéticos que diz respeito a um programa de trabalho do governo que tenta colocar pessoas desfavorecidas em empregos melhores. Uma amostra dos participantes potenciais do programa foi seguida, alguns dos quais foram selecionados aleatoriamente para inscrição no programa, enquanto outros não foram. Cada caso representa um participante separado do programa.
- **worldsales.sav** Este arquivo de dados hipotéticos contém a receita de vendas por continente e produto.

19. Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, e B. G. Ferris Jr., 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366-374.

Avisos

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

LICENÇA DE COPYRIGHT:

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

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

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

Marcas comerciais

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

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

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

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

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

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

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

Índice Remissivo

A

- aninhando variáveis
 - Tabelas customizadas 27, 28
 - variáveis de escala 53
- arquivos de amostra
 - posição 80

C

- casas decimais
 - controlando o número de casas decimais exibidas em tabelas customizadas 6
 - controlando o número de decimais exibidos em tabelas customizadas 23, 76
- categorias calculadas
 - a partir de subtotaís 39
 - formatos de exibição 15
 - ocultando categorias na expressão 38
 - Tabelas customizadas 14, 37
- categorias pós-calculadas
 - Tabelas customizadas 14, 37
- células vazias
 - valor exibido em tabelas customizadas 16, 78
- conjuntos de múltiplas respostas 66
 - porcentagens 9
 - respostas duplicadas em conjuntos de categorias múltiplas 16
 - teste de significância 66, 72
- contagem
 - versus N válido 51
- controlando o número de casas decimais exibidas 23

D

- data
 - incluindo a data atual em tabelas customizadas 17
- desvio padrão
 - Tabelas customizadas 9

E

- empilhando variáveis
 - diversas variáveis de origem
 - estatística de sumarização 47
 - empilhando variáveis de estrato 31
 - estatísticas básicas diferentes para variáveis diferentes 52
 - Tabelas customizadas 26
 - variáveis de escala 50
- estatísticas
 - estatísticas básicas 44
 - estatísticas básicas total customizadas 47
 - tabelas empilhadas 47

- estatísticas básicas 44
 - dimensão de origem 45
 - estatísticas básicas total customizadas 47
 - formato de exibição 76
 - mudando o texto do rótulo 77
 - sumarizações diferentes para variáveis diferentes em tabelas empilhadas 52
 - tabelas empilhadas 47
 - variável de origem 45
- estatísticas básicas diferentes para variáveis diferentes
 - tabelas empilhadas 52
- estatísticas básicas total customizadas 47
- estatísticas de médias de coluna
 - tabelas customizadas 58
- estatísticas de proporção da coluna
 - tabelas customizadas 61
- estatísticas de teste
 - Tabelas customizadas 18, 55
- excluindo categorias
 - Tabelas customizadas 12, 24
- exibindo valores de categoria 48

F

- formatos de exibição 23
 - estatísticas básicas em tabelas customizadas 11, 76

H

- horário
 - incluindo o horário atual em tabelas customizadas 17

I

- imprimindo tabelas com estratos 31
- intervalo
 - Tabelas customizadas 9
- intervalos de confiança 54

L

- largura da coluna
 - controlando em tabelas customizadas 16, 78
- legendas
 - Tabelas customizadas 17

M

- máximo
 - Tabelas customizadas 9
- média 50
 - Tabelas customizadas 9
- mediana 50

- mediana (*continuação*)
 - Tabelas customizadas 9
- mínimo
 - Tabelas customizadas 9
- modo
 - Tabelas customizadas 9

N

- N total 74
- N válido 51, 74
 - Tabelas customizadas 9
- nível de medição
 - mudando em tabelas customizadas 1

O

- ocultando rótulos de estatísticas em tabelas customizadas 20
- omitindo categorias
 - Tabelas customizadas 24
- ordenando categorias
 - Tabelas customizadas 24

P

- porcentagens
 - conjuntos de múltiplas respostas 9
 - em tabelas customizadas 7, 8, 21, 22
 - valores omissos 74
- processamento de arquivo dividido
 - tabelas customizadas 5

Q

- qui-quadrado
 - Tabelas customizadas 55

R

- reduzindo categorias
 - Tabelas customizadas 36
- reordenando categorias
 - Tabelas customizadas 12
- rótulos
 - mudando o texto do rótulo para estatísticas básicas 77
- rótulos de canto
 - Tabelas customizadas 17
- rótulos de variáveis
 - suprimindo a exibição em tabelas customizadas 6

S

- soma
 - Tabelas customizadas 9
- subtotais 35

subtotais (*continuação*)
ocultando o subtotal das
categorias 36
Tabelas customizadas 12, 32
sumarizações agrupadas
variáveis de escala 52

T

tabela de frequências
Tabelas customizadas 15, 41
tabelas
Tabelas customizadas 1
tabelas comperimeter 15, 41
tabelas customizadas
processamento de arquivo dividido 5
Tabelas customizadas
aninhando variáveis 27, 28
aninhando variáveis de estrato 31
categorias calculadas 12, 14, 37
categorias pós-calculadas 14, 37
células vazias 16
como construir uma tabela 4
conjuntos de categorias múltiplas 16
conjuntos de múltiplas respostas 1,
66
controlando o número de casas
decimais exibidas 6
dimensão de origem estatística 22
empilhando variáveis 26
estatísticas básicas 7, 8, 9
estatísticas de teste 18, 55
excluindo categorias 12, 24
exclusão de valores omissos para
sumarizações de escala 16
formatos de exibição 6
formatos de exibição de estatísticas
básicas 11
imprimindo tabelas em estratos 31
largura da coluna 16
legendas 17
mostrando e ocultando nomes e
rótulos de variáveis 6
mudando a dimensão de estatísticas
básicas 10
mudando o nível de medição 1
ocultando o subtotal das
categorias 36
ocultando rótulos de estatísticas 20
ordenando categorias 24
porcentagens 7, 8, 21, 22
porcentagens de linhas versus da
coluna 21
porcentagens para conjuntos de
múltiplas respostas 9
reduzindo categorias 36
reordenando categorias 12
rótulos de canto 17
rótulos de valor para variáveis
categóricas 1
subtotais 12, 32
tabela de frequências 15, 41
tabelas comperimeter 15, 41
tabelas de frequência de média 10
tabelas de variáveis com categorias
compartilhadas 15, 41

Tabelas customizadas (*continuação*)
tabelas simples para variáveis
categóricas 20
tabulação cruzada 22
teste de significância e múltiplas
respostas 66
títulos 17
totais 12, 21, 32
totais customizados 10
totais em tabelas com categorias
excluídas 24
totais marginais 23
trocando variáveis de linha e de
coluna 29
uma amplitude relativamente
pequena 21
variáveis categóricas 1
variáveis de escala 1
variáveis de estrato 30, 31
visualização compacta 28
tabelas de frequência de média 10, 47
tabulação cruzada
Tabelas customizadas 22
testes de significância
conjuntos de múltiplas respostas 72
Tabelas customizadas 18
títulos
Tabelas customizadas 17
totais
categorias excluídas 32
estratos 35
posição de exibição 33
Tabelas aninhadas 33
Tabelas customizadas 12, 21, 32
totais de grupo 33
totais marginais para tabelas
customizadas 23
totais de grupo 33
totais de subgrupo 33

V

valores
exibindo rótulo e valores de
categoria 48
valores e rótulos de valor 48
valores omissos 51, 73
efeito sobre os cálculos de
porcentagem 74
incluindo em tabelas
customizadas 74
valores omissos de usuário 73
valores omissos do sistema 73
variância
Tabelas customizadas 9
variáveis de escala
aninhadas 53
empilhando 50
estatísticas básicas 50
sumarizações agrupadas 52
sumarizações agrupadas por variáveis
categóricas de linha e de coluna 53
várias estatísticas básicas 50
variáveis de estrato
aninhando variáveis de estrato 31
empilhando variáveis de estrato 31
imprimindo tabelas em estratos 31

variáveis de estrato (*continuação*)
Tabelas customizadas 30, 31
variável de origem estatística de
sumarização
variáveis de escala 53



Impresso no Brasil