

*Guia rápido do IBM SPSS Statistics 28*



**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 28, liberação 0, modificação 0 do IBM® SPSS Statistics e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

© **Copyright International Business Machines Corporation .**

---

# Índice

<b>Capítulo 1. Introdução.....</b>	<b>1</b>
Arquivos de Amostra.....	1
Abrindo um arquivo de dados.....	1
Executando uma análise .....	2
Criação de gráficos.....	4
<b>Capítulo 2. Lendo dados.....</b>	<b>7</b>
Estrutura básica de arquivos de dados do IBM SPSS Statistics.....	7
Lendo arquivos de dados do IBM SPSS Statistics.....	7
Lendo dados do Excel.....	8
Lendo dados de um banco de dados.....	11
Lendo dados de um arquivo de texto.....	14
<b>Capítulo 3. Usando o Editor de Dados.....</b>	<b>19</b>
Inserindo dados numéricos.....	19
Inserindo dados de sequência.....	20
Definindo dados.....	21
Incluindo rótulos de variáveis.....	21
Alterando o formato e o tipo de variável.....	22
Incluindo rótulos de valor.....	22
Manipulando dados omissos.....	23
Valores omissos para uma variável numérica.....	23
Valores omissos para uma variável de sequência.....	24
<b>Capítulo 4. Examinando estatísticas de resumo para variáveis individuais.....</b>	<b>25</b>
Nível de medição.....	25
Medidas de resumo para dados categóricos.....	25
Gráficos para dados categóricos.....	26
Medidas de resumo para variáveis de escala.....	27
Histogramas para variáveis de escala.....	28
<b>Capítulo 5. Criando e editando gráficos.....</b>	<b>31</b>
Fundamentos para criação de gráfico.....	31
Usando a galeria do Construtor de Gráfico.....	31
Definindo variáveis e estatísticas.....	32
Incluindo Texto.....	33
Criando o gráfico.....	34
<b>Capítulo 6. Trabalhando com a Saída.....</b>	<b>37</b>
Usando o Visualizador.....	37
Usando o Editor de Tabela Dinâmica.....	38
Acessando definições de saída.....	38
Tabelas dinâmicas.....	39
Criando e exibindo estratos.....	40
Editando tabelas.....	40
Ocultando linhas e colunas.....	41
Alterando formatos de exibição de dados.....	42
TableLooks.....	43
Usando formatos predefinidos.....	43
Customizando estilos do TableLook.....	44

Mudando os formatos de tabela padrão.....	46
Customizando as configurações de exibição iniciais.....	46
Exibindo rótulos de variável e de valor.....	47
Usando resultados em outros aplicativos.....	48
Colando resultados como tabelas do Word.....	49
Colando resultados como texto.....	49
Exportando resultados para arquivos do Microsoft Word, PowerPoint e Excel.....	50
Exportando resultados para PDF.....	55
Exportando resultados para HTML.....	57
<b>Capítulo 7. Trabalhando com sintaxe.....</b>	<b>59</b>
Colando sintaxe.....	59
Editando a sintaxe.....	60
Abrindo e executando um arquivo de sintaxe.....	61
Usando pontos de interrupção.....	61
<b>Capítulo 8. Modificando valores de dados.....</b>	<b>63</b>
Criando uma variável categórica a partir de uma variável de escala.....	63
Calculando novas variáveis.....	65
Usando funções em expressões.....	66
Usando expressões condicionais.....	67
Trabalhando com datas e horas.....	67
Calculando o período de tempo entre duas datas.....	68
Incluindo uma duração para uma data.....	69
<b>Capítulo 9. Classificando e selecionando dados.....</b>	<b>71</b>
Classificando Dados.....	71
Processamento de divisão de arquivo.....	71
Classificando casos para o processamento de divisão de arquivo.....	73
Ligando e desligando o processamento de divisão de arquivo.....	73
Selecionando subconjuntos de casos.....	73
Selecionando casos com base em expressões condicionais.....	74
Selecionando uma amostra aleatória.....	75
Selecionando um intervalo de tempo ou um intervalo de caso.....	76
Tratamento de casos não selecionados.....	76
Status de seleção de caso.....	77
<b>Capítulo 10. Arquivos de Amostra.....</b>	<b>79</b>
<b>Avisos.....</b>	<b>89</b>
Marcas comerciais.....	90
<b>Índice Remissivo.....</b>	<b>93</b>

# Capítulo 1. Introdução

Este guia irá mostrar como usar muitos dos recursos disponíveis. Ele é projetado para fornecer um guia passo-a-passo, prático. Todos os arquivos apresentados nos exemplos são instalados com o aplicativo para que seja possível acompanhar, executando as mesmas análises e obtendo os mesmos resultados mostrados aqui.

Se quiser exemplos detalhados de várias técnicas de análise estatística, tente os Estudos de Casos passo a passo, disponíveis a partir do menu de Ajuda.

## Arquivos de Amostra

A maioria dos exemplos apresentados aqui usa o arquivo de dados *demo.sav*. Esse arquivo de dados é uma pesquisa fictícia de vários milhares de pessoas, contendo informações básicas demográficas e do consumidor.

Se você estiver usando a versão Student, sua versão de *demo.sav* será uma amostra representativa do arquivo de dados original, reduzido para atender ao limite de 1.500 casos. Os resultados que você obtém usando esse arquivo de dados serão diferentes dos resultados mostrados aqui.

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.

## Abrindo um arquivo de dados

Para abrir um arquivo de dados:

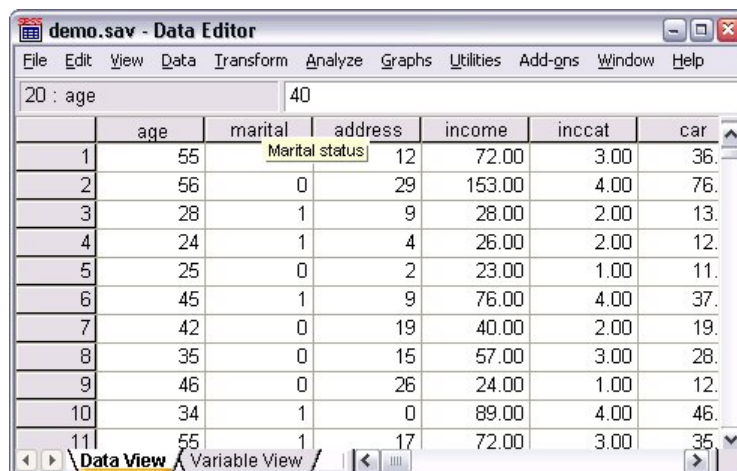
1. Nos menus, escolha:

**Arquivo > Abrir > Dados...**

Uma caixa de diálogo para abrir arquivos é exibida.

Por padrão, os arquivos de dados do IBM SPSS Statistics (extensão *.sav*) são exibidos.

Este exemplo usa o arquivo *demo.sav*.



	age	marital	address	income	inccat	car
1	55	12	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

Figura 1. Arquivo *demo.sav* no Editor de Dados

O arquivo de dados é exibido no Editor de Dados. Em Visualização de Dados, se você colocar o cursor do mouse sobre um nome de variável (os títulos das colunas), um rótulo de variável mais descritivo será exibido (se um rótulo tiver sido definido para essa variável).

Por padrão, os valores de dados reais são exibidos. Para exibir rótulos:

2. Nos menus, escolha:

**Visualizar > Rótulos de valor**



Figura 2. Botão Rótulos de Valor

Como alternativa, você pode usar o botão Rótulos de Valor na barra de ferramentas.

A captura de tela mostra a janela 'demo.sav - Data Editor' do SPSS. O menu 'Visualizar' está aberto, e a opção 'Rótulos de valor' está selecionada. A tabela de dados abaixo mostra os resultados com rótulos descritivos para as variáveis 'income' e 'inccat'.

	age	marital	address	income	inccat	car
1	55	Married	12	72.00	\$50 - \$74	36.
2	56	Unmarried	29	153.00	\$75+	76.
3	28	Married	9	28.00	\$25 - \$49	13.
4	24	Married	4	26.00	\$25 - \$49	12.
5	25	Unmarried	2	23.00	Under \$25	11.
6	45	Married	9	76.00	\$75+	37.
7	42	Unmarried	19	40.00	\$25 - \$49	19.
8	35	Unmarried	15	57.00	\$50 - \$74	28.
9	46	Unmarried	26	24.00	Under \$25	12.
10	34	Married	0	89.00	\$75+	46.
11	55	Married	17	72.00	\$50 - \$74	35.

Figura 3. Rótulos de valor exibidos no Editor de Dados

Os rótulos de valor descritivo são, agora, exibidos para facilitar a interpretação das respostas.

## Executando uma análise

Se você tiver alguma opção complementar, o menu Analisar conterá uma lista das categorias de análise estatística e de relatório.

Vamos começar criando uma tabela de frequências simples (tabela de contagens). Este exemplo requer a opção Base de Estatísticas.

1. Nos menus, escolha:

**Analisar > Estatísticas Descritivas > Frequências...**

A caixa de diálogo Frequências é exibida.

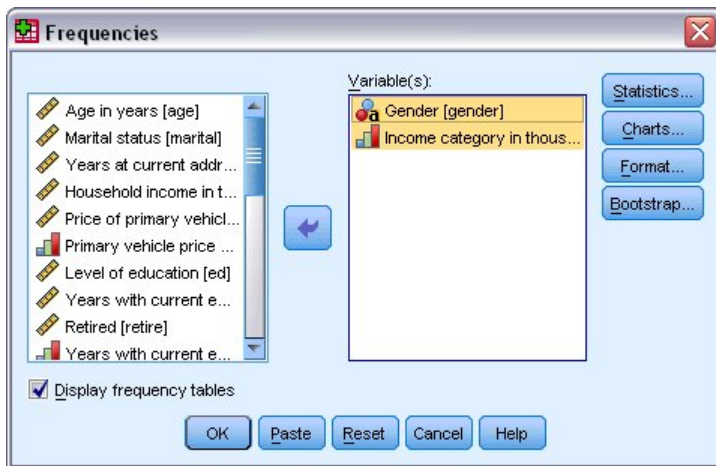


Figura 4. Caixa de diálogo Frequências

Um ícone próximo a cada variável fornece informações sobre o tipo de dados e o nível de medição.

	Numérico	Sequência	Date	Hora
Escala (Contínua)		n/a		
Ordinal				
Nominal				

Se o rótulo de variável e/ou o nome aparecer truncado na lista, o rótulo/nome completo será exibido quando o cursor for posicionado sobre ele. O nome de variável *inccat* é exibido entre colchetes depois do rótulo de variável descritivo. *Categoria de receita em milhares* é o rótulo de variável. Se não houvesse nenhum rótulo de variável, apenas o nome de variável poderia aparecer na caixa de listagem.

Você pode redimensionar as caixas de diálogo exatamente como as janelas, clicando e arrastando as bordas ou os cantos externos. Por exemplo, se você aumentar a caixa de diálogo, as listas de variáveis também serão maiores.

Na caixa de diálogo, você escolhe as variáveis que deseja analisar na lista de origem à esquerda e as arrasta e solta na lista Variável(is) à direita. O botão **OK**, que executa a análise, é desativado até pelo menos uma variável ser colocada na lista Variável(is).

Em muitos diálogos, você pode obter informações adicionais clicando com o botão direito em qualquer nome de variável na lista e selecionando **Informações de Variáveis** no menu pop-up.

2. Clique em *Sexo [sexo]* na lista de variáveis de origem e arraste a variável para a lista Variável(is) de destino.
3. Clique em *Categoria de receita em milhares [inccat]* na lista de origem e arraste-a para a lista de destino.

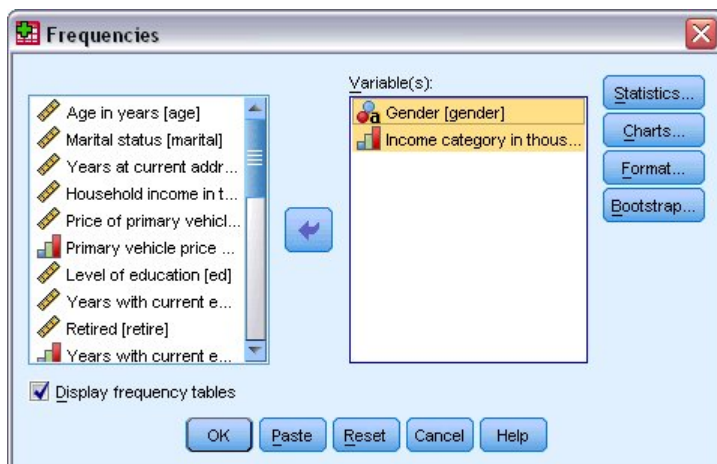


Figura 5. Variáveis selecionadas para análise

4. Clique em **OK** para executar o procedimento.

Os resultados são exibidos na janela Visualizador.

Gender					
		Frequency	Percent	Valid Percent	Cumul: Perce
Valid	Female	3179	49.7	49.7	
	Male	3221	50.3	50.3	
	Total	6400	100.0	100.0	

Income category in thousands					
		Frequency	Percent	Valid Percent	Cum Pe
Valid	Under \$25	1174	18.3	18.3	
	\$25 - \$49	2388	37.3	37.3	
	\$50 - \$74	1120	17.5	17.5	
	\$75+	1718	26.8	26.8	
	Total	6400	100.0	100.0	

Figura 6. Tabela de frequência de categorias de receita

## Criação de gráficos

Embora alguns procedimentos estatísticos possam criar gráficos, você também pode usar o menu Gráficos para criar gráficos.

Por exemplo, você pode criar um gráfico que mostre o relacionamento entre o serviço telefônico wireless e a propriedade PDA (Personal Digital Assistant).

1. Nos menus, escolha:

**Gráficos > Construtor de Gráfico...**



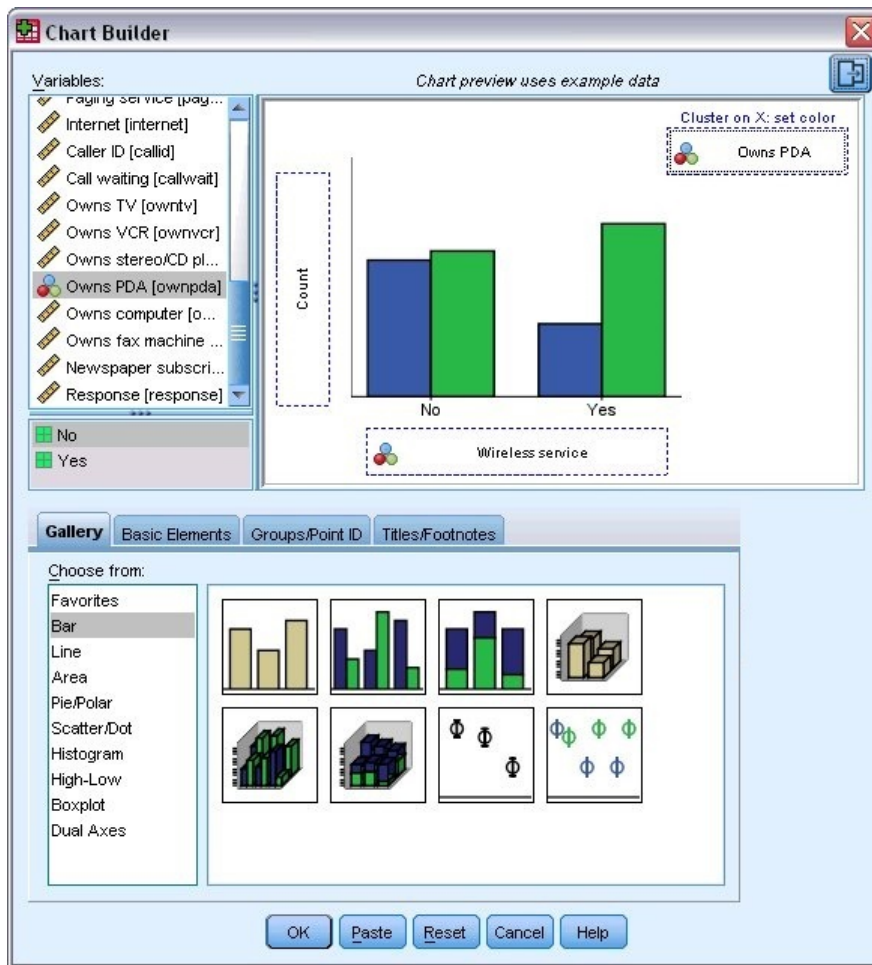


Figura 7. Caixa de diálogo Construtor de Gráfico com áreas de lançamento completas

2. Clique na guia **Galeria** (se ela não estiver selecionada).
3. Clique em **Barra** (se esta opção não estiver selecionada).
4. Arraste o ícone Barra em Cluster para a tela, que é a área grande acima da Galeria.
5. Role para baixo a lista Variáveis, clique com o botão direito em *Serviço wireless [wireless]* e, em seguida, escolha **Nominal** como o nível de medição.
6. Arraste a variável *Serviço wireless [wireless]* para o eixo x.
7. Clique com o botão direito em *Possui PDA [ownpda]* e escolha **Nominal** como o nível de medição.
8. Arraste a variável *Possui PDA [ownpda]* para a área de lançamento de cluster no canto superior direito da tela.
9. Clique em **OK** para criar o gráfico.

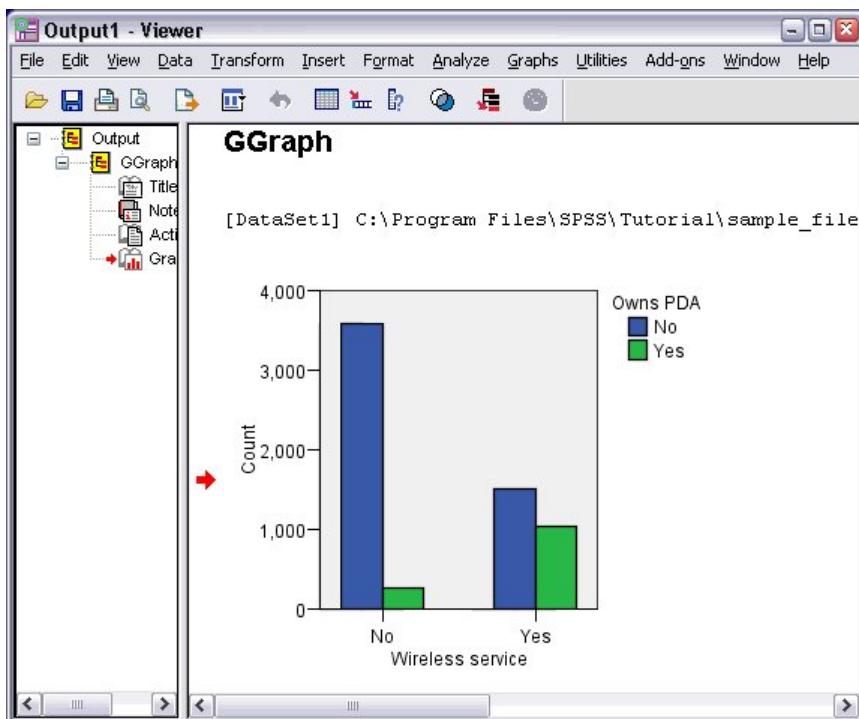


Figura 8. Gráfico de barras exibido na janela Visualizador

O gráfico de barras é exibido no Visualizador. O gráfico mostra que pessoas com o serviço telefônico wireless muito mais provavelmente terão PDAs do que as pessoas sem o serviço wireless.

É possível editar gráficos e tabelas clicando duas vezes neles na área de janela de conteúdo da janela Visualizador e é possível copiar e colar os resultados em outros aplicativos. Esses tópicos serão abrangidos mais tarde.

---

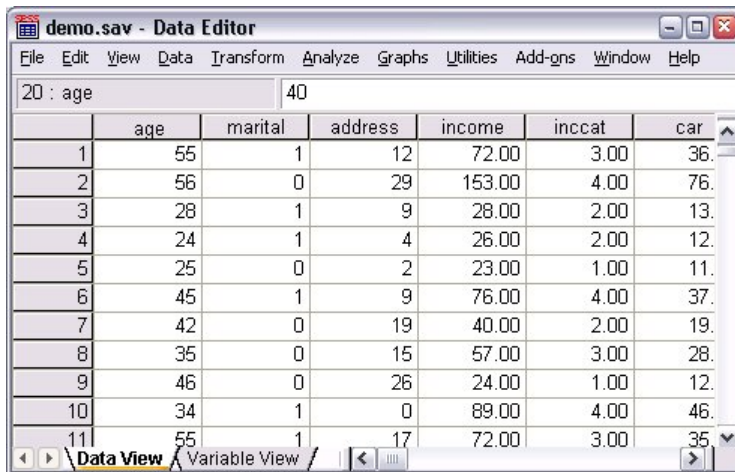
## Capítulo 2. Lendo dados

Os dados podem ser inseridos diretamente ou podem ser importados de várias origens diferentes. Os processos para a leitura de dados armazenados nos arquivos de dados do IBM SPSS Statistics; aplicativos de planilha, como Microsoft Excel; aplicativos de banco de dados, como Microsoft Access; e arquivos de texto são todos discutidos neste capítulo.

---

### Estrutura básica de arquivos de dados do IBM SPSS Statistics

---



The screenshot shows the 'demo.sav - Data Editor' window. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The current view is 'Data View'. The data table has the following structure:

	age	marital	address	income	inccat	car	
1	55	1	12	72.00	3.00	36.	
2	56	0	29	153.00	4.00	76.	
3	28	1	9	28.00	2.00	13.	
4	24	1	4	26.00	2.00	12.	
5	25	0	2	23.00	1.00	11.	
6	45	1	9	76.00	4.00	37.	
7	42	0	19	40.00	2.00	19.	
8	35	0	15	57.00	3.00	28.	
9	46	0	26	24.00	1.00	12.	
10	34	1	0	89.00	4.00	46.	
11	55	1	17	72.00	3.00	35.	

Figura 9. Editor de Dados

Os arquivos de dados do IBM SPSS Statistics são organizados por casos (linhas) e variáveis (colunas). Neste arquivo de dados, os casos representam entrevistados individuais para uma pesquisa. As variáveis representam respostas a cada pergunta feita na pesquisa.

---

### Lendo arquivos de dados do IBM SPSS Statistics

---

Os arquivos de dados do IBM SPSS Statistics, que têm uma extensão de arquivo *.sav*, contêm os dados salvos.

1. Nos menus, escolha:

**Arquivo > Abrir > Dados...**

2. Procure e abra *demo.sav*. Consulte o tópico [Capítulo 10, “Arquivos de Amostra”](#), na página 79 para obter mais informações

Os dados são, agora, exibidos no Editor de Dados.

The screenshot shows the SPSS Data Editor window for a file named 'demo.sav'. The window title bar includes the SPSS logo and the text 'demo.sav - Data Editor'. Below the title bar is a menu bar with the following options: File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The main area of the window displays a data table with 11 rows and 6 columns. The columns are labeled 'age', 'marital', 'address', 'income', 'inccat', and 'car'. The rows contain numerical data. At the top of the table, there is a row with the label '20 : age' and the value '40'. At the bottom of the table, there is a row with the label '11' and the value '55'. The status bar at the bottom of the window shows 'Data View' and 'Variable View'.

	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

Figura 10. Arquivo de dados aberto

## Lendo dados do Excel

Em vez de digitar todos os seus dados diretamente no Editor de Dados, você pode ler dados a partir de aplicativos como o Microsoft Excel. É possível também ler os títulos de colunas como nomes de variáveis.

1. Nos menus, escolha:

**Arquivo > Importar dados > Excel**

2. Acesse a pasta Amostras\Inglês e selecione demo.xlsx.

O diálogo **Ler Arquivo do Excel** exibe uma visualização do arquivo de dados. O conteúdo da primeira planilha no arquivo é exibido. Se o arquivo tiver várias planilhas, você poderá selecionar a planilha na lista.

É possível ver que alguns dos valores de sequência para *Sexo* têm espaços extras. Alguns dos valores para *MaritalStatus* são exibidos como pontos (.).

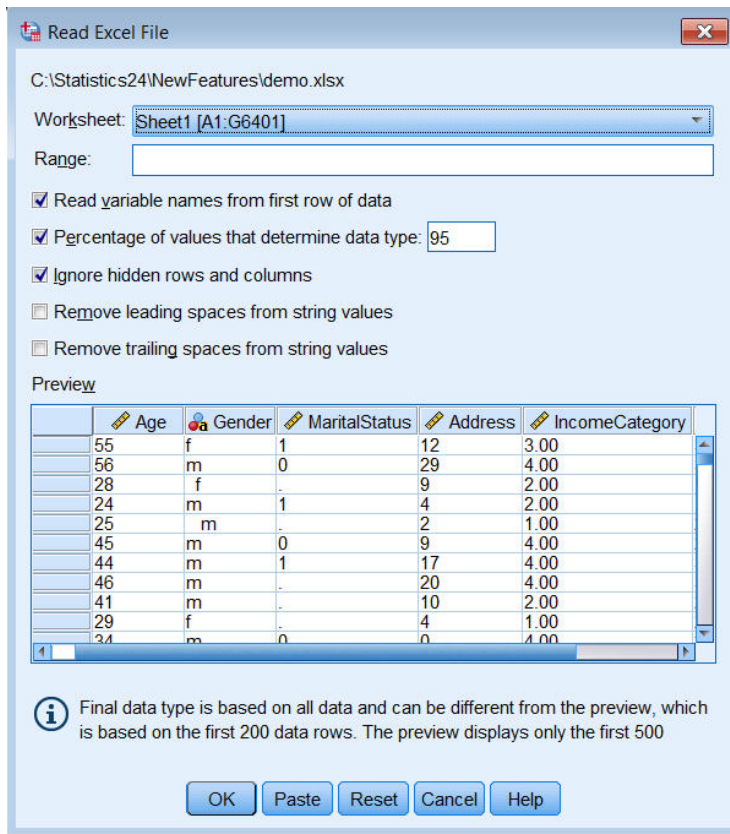
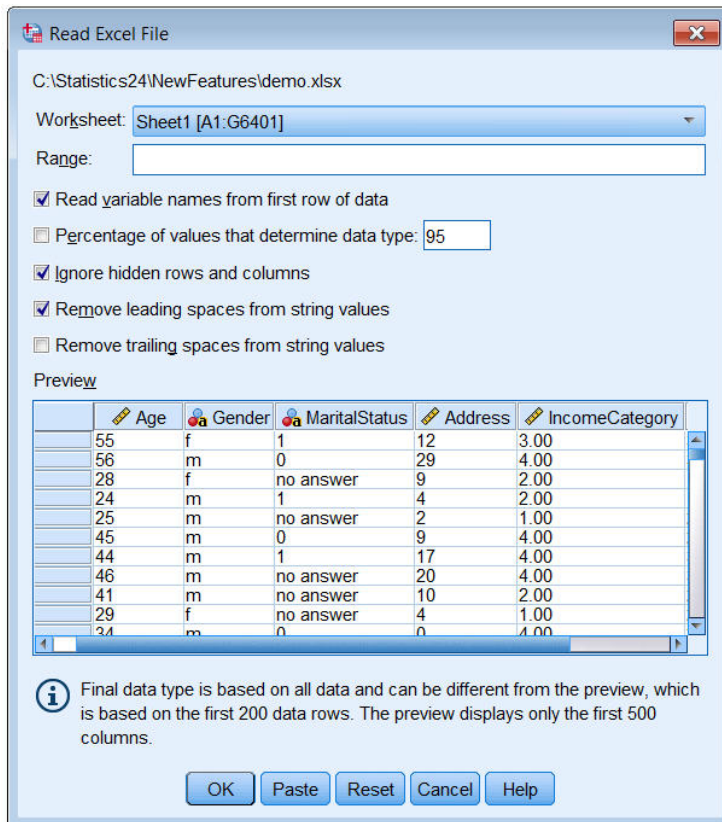


Figura 11. Diálogo Ler arquivo do Excel

3. Certifique-se de que a opção **Ler nomes de variáveis a partir da primeira linha de dados** seja selecionada. Se os títulos das colunas não estiverem em conformidade com as regras de nomes de variáveis, eles serão convertidos em nomes de variáveis válidos. Os títulos de colunas originais são salvos como rótulos de variáveis.
4. Selecione **Remover espaços extras de valores de sequência**.
5. Cancele a seleção de **Porcentagem de valores que determinam o tipo de dados**.



O valor de sequência "nenhuma resposta" agora é exibido nas células que estavam omissas do sistema. Se não houver nenhuma porcentagem do parâmetro de valores e a coluna contiver uma combinação de tipo de dados, a variável será lida como um tipo de dados de sequência. Todos os valores são preservados, mas os valores numéricos são tratados como valores de sequência.

6. Selecione (marque) **Porcentagem de valores que determinam o tipo de dados** para tratar *MaritalStatus* como uma variável numérica.
7. Clique em **OK** para ler o arquivo do Excel.

Agora, os dados aparecem no Editor de Dados, com os títulos de colunas usados como nomes de variáveis. Como os nomes de variáveis não podem conter espaços, os espaços dos títulos de colunas originais são removidos. Por exemplo, o título da coluna "Estado Civil" é convertido na variável *MaritalStatus*. O título de coluna original é retido como um rótulo de variável.

\*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help

Visible: 6 of 6 Variables

	Age	Gender	MaritalStatus	Address	IncomeCategory	JobCategory
1	55 f		1	12	3.00	3
2	56 m		0	29	4.00	3
3	28 f		.	9	2.00	1
4	24 m		1	4	2.00	1
5	25 m		.	2	1.00	2
6	45 m		0	9	4.00	2
7	44 m		1	17	4.00	3
8	46 m		.	20	4.00	3
9	41 m		.	10	2.00	2
10	29 f		.	4	1.00	2

Data View Variable View

Figura 12. Dados do Excel importados

### Informações relacionadas

“Arquivos de Amostra” na página 79

## Lendo dados de um banco de dados

Dados de origens de banco de dados são facilmente importados usando o Assistente de Base de Dados. Qualquer banco de dados que use drivers ODBC (Open Database Connectivity) pode ser lido diretamente após os drivers serem instalados. Os drivers ODBC para vários formatos de banco de dados são fornecidos no CD de instalação. Drivers adicionais podem ser obtidos de outros fornecedores. Um dos aplicativos de banco de dados mais comuns, Microsoft Access, é discutido neste exemplo.

*Nota:* Este exemplo é específico para o Microsoft Windows e requer um driver ODBC para Access. O driver ODBC do Microsoft Access funciona apenas com a versão de 32 bits do IBM SPSS Statistics. As etapas são semelhantes em outras plataformas, mas podem requerer um driver ODBC de terceiro para Access.

1. Nos menus, escolha:

**Arquivo > Importar dados > Banco de dados > Nova consulta...**

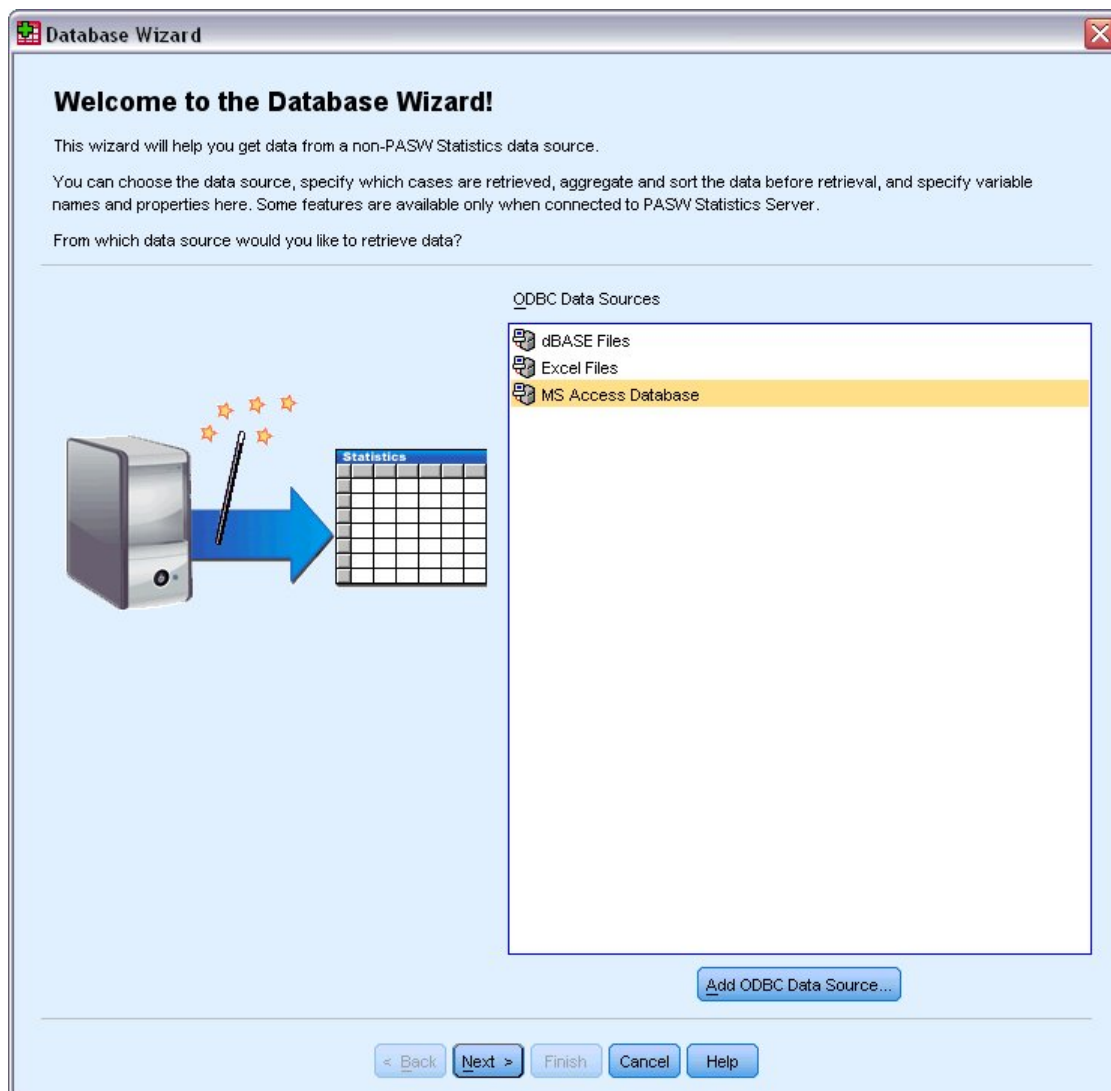


Figura 13. Caixa de diálogo Bem-vindo do Assistente de Base de Dados

2. Selecione **Banco de Dados do MS Access** na lista de origens de dados e clique em **Avançar**.

*Nota:* Dependendo de sua instalação, você também pode ver uma lista de origens de dados OLEDB no lado esquerdo do assistente (apenas sistemas operacionais Windows), mas este exemplo usa a lista de origens de dados ODBC exibida no lado direito.

3. Clique em **Procurar** para navegar para o arquivo de banco de dados do Access que você deseja abrir.
4. Abra *demo.mdb*. Consulte o tópico [Capítulo 10, “Arquivos de Amostra”](#), na página 79 para obter mais informações
5. Clique em **OK** na caixa de diálogo de login.

Na próxima etapa, você pode especificar as tabelas e variáveis que deseja importar.



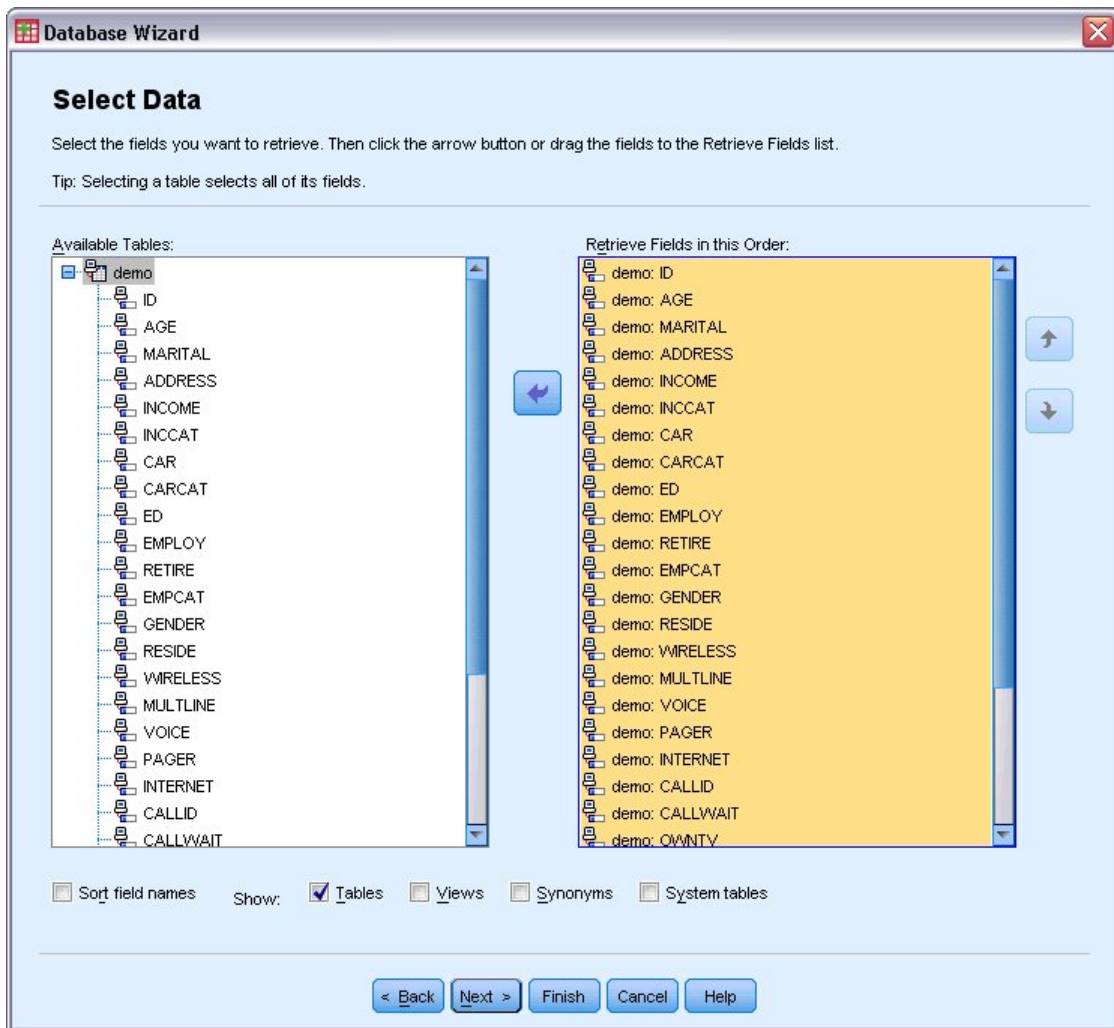


Figura 14. Etapa Selecionar Dados

6. Arraste a tabela **demo** inteira para a lista Recuperar Campos nesta Ordem.
7. Clique em **Avançar**.

Na próxima etapa, você pode selecionar quais registros (casos) importar.

Se não quiser importar todos os casos, você poderá importar um subconjunto de casos (por exemplo, homens com mais de 30) ou poderá importar uma amostra aleatória de casos da origem de dados. Para origens de dados grandes, você pode querer limitar o número de casos a uma amostra pequena, representativa para reduzir o tempo de processamento.

8. Clique em **Avançar** para continuar.

Os nomes de campos são usados para criar nomes de variáveis. Se necessário, os nomes são convertidos em nomes de variáveis válidos. Os nomes de campos originais são preservados como rótulos de variáveis. É possível também alterar os nomes de variáveis antes de importar o banco de dados.

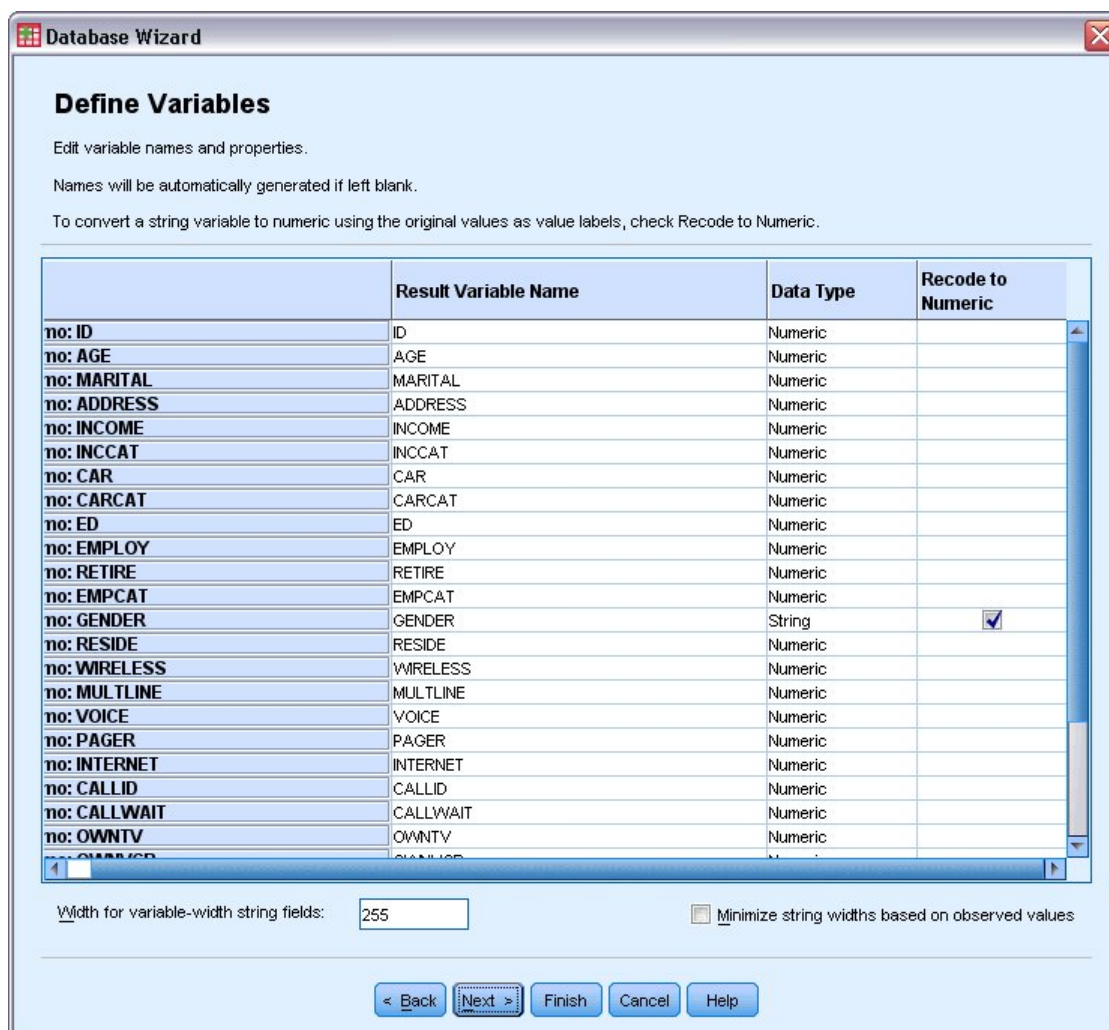


Figura 15. Etapa Definir Variáveis

9. Clique na célula **Recodificar para Numérico** no campo Sexo. Esta opção converte variáveis de sequência em variáveis de número inteiro e retém o valor original como o rótulo de valor para a nova variável.
10. Clique em **Avançar** para continuar.

A instrução SQL criada a partir de suas seleções no Assistente de Base de Dados aparece na etapa Resultados. Esta instrução pode ser executada agora ou salva em um arquivo para uso posterior.

11. Clique em **Concluir** para importar os dados.

Todos os dados no banco de dados do Access que você selecionou para importação estão agora disponíveis no Editor de Dados.

## Lendo dados de um arquivo de texto

Arquivos de texto representam outra origem comum de dados. Muitos programas de planilha e bancos de dados podem salvar seu conteúdo em um dos muitos formatos de arquivos de texto. Arquivos delimitados por vírgula ou por tabulação referem-se a linhas de dados que usam vírgulas ou tabulações para indicar cada variável. Neste exemplo, os dados são delimitados por tabulação.

1. Nos menus, escolha:

**Arquivo > Importar dados > Dados de texto**

2. Acesse a pasta Amostras\Inglês e selecione demo.txt.

O Assistente de Importação de Texto o orienta através do processo de definição de como o arquivo de texto especificado é interpretado.

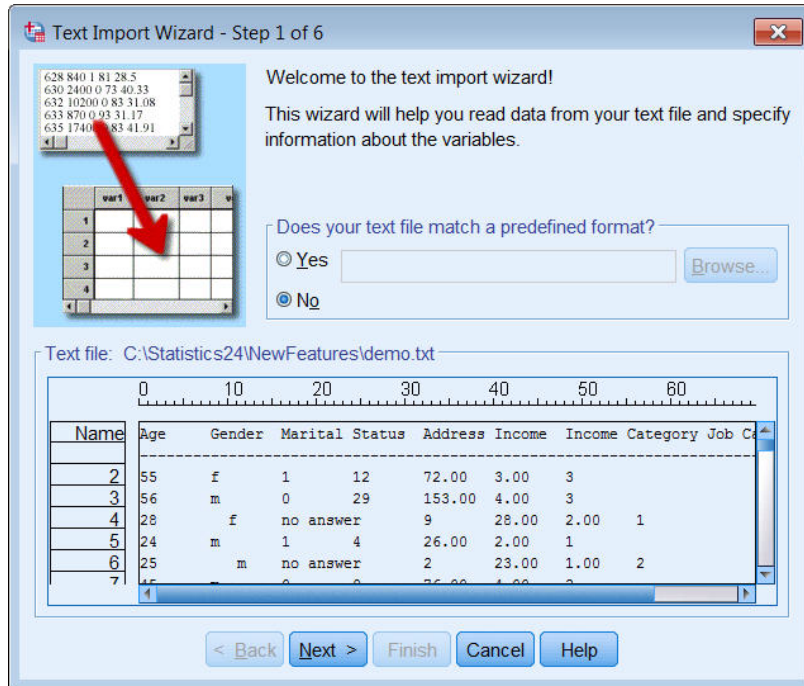


Figura 16. Assistente de Importação de Texto: Etapa 1 de 6

3. Na Etapa 1, é possível escolher um formato predefinido ou criar um novo formato no assistente. Selecione **Não**.

4. Clique em **Avançar** para continuar.

Conforme indicado anteriormente, este arquivo usa a formatação delimitada por tabulação. Além disso, os nomes de variáveis são definidos na linha superior deste arquivo.

5. Na etapa 2 do assistente, selecione **Delimitado** para indicar que o arquivo tem uma estrutura de formatação delimitada.

6. Selecione **Sim** para indicar que o arquivo inclui nomes de variáveis na parte superior do arquivo.

7. Clique em **Avançar** para continuar.

8. Na etapa 3, insira 2 para o número da linha em que começa o primeiro caso de dados (porque os nomes de variáveis estão na primeira linha).

9. Mantenha os valores padrão para o restante desta etapa e clique em **Avançar** para continuar.

A Visualização de dados na Etapa 4 fornece a você uma maneira rápida de assegurar que o arquivo seja lido corretamente

10. Selecione **Tab** e cancele a seleção das outras opções para delimitadores. **Espaço** é selecionado por padrão porque o arquivo contém espaços. Para este arquivo, os espaços fazem parte dos valores de dados, não delimitadores. É necessário cancelar a seleção de **Espaço** para ler o arquivo corretamente.

11. Selecione **Remover espaços extras para valores de sequência**. Os espaços no início dos valores de sequência afetam como os valores de sequência são avaliados em expressões. Neste arquivo, alguns valores para Sexo têm espaços extras que não fazem parte do valor. Se você não remover esses espaços, um valor de " f" será tratado como um valor diferente de "f".

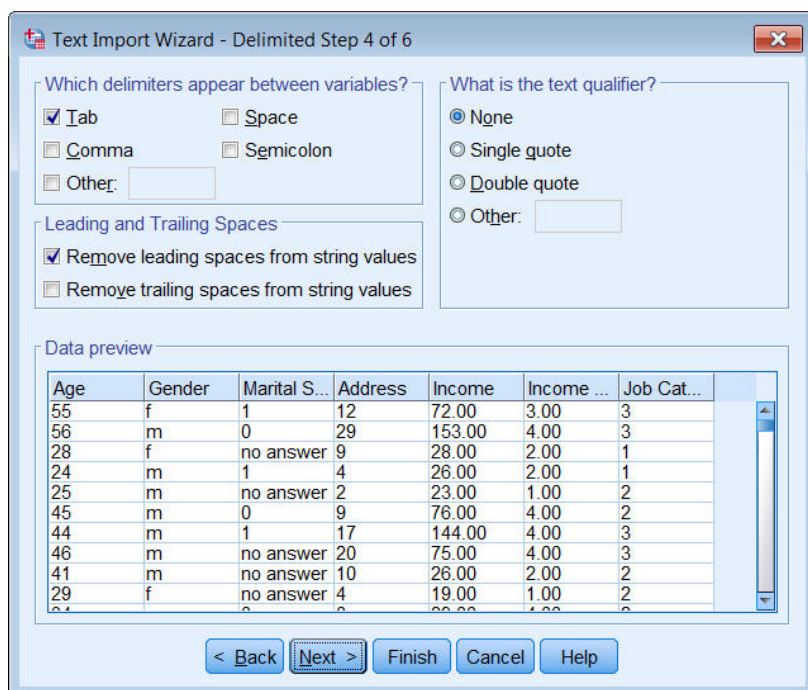


Figura 17. Assistente de Importação de Texto: Etapa 4 de 6

12. Clique em **Avançar** para continuar.

Como os nomes de variáveis são modificados para que fiquem em conformidade com as regras de nomenclatura, a etapa 5 dá a você a oportunidade de editar quaisquer nomes indesejáveis.

Os tipos de dados podem ser definidos aqui também. Por exemplo, é possível alterar *Receita* para o formato da moeda de dólar.

Para alterar um tipo de dados:

13. Na **Visualização de dados**, selecione *Receita*.
14. Selecione **Dólar** na lista suspensa Formato de Dados.

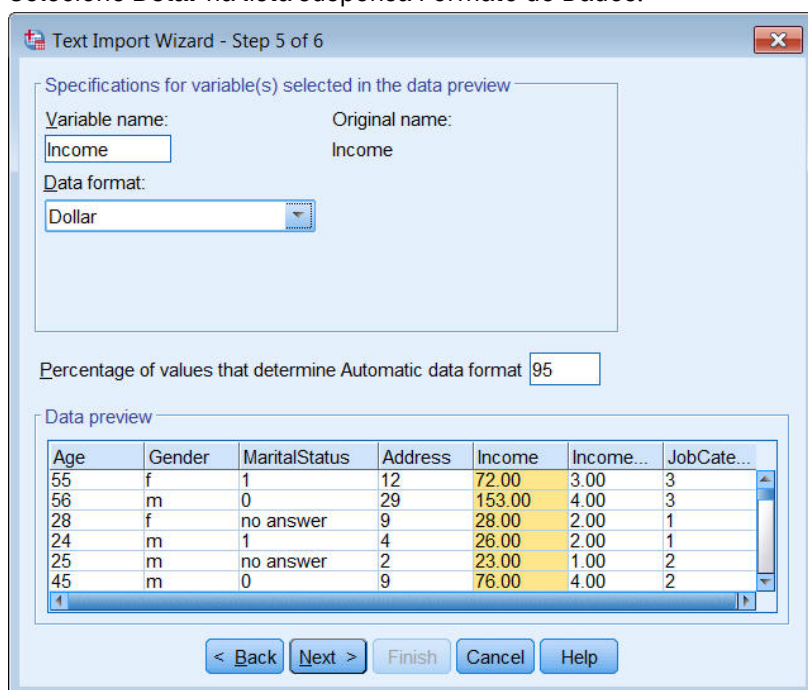


Figura 18. Mudar o tipo de dados

A variável *MaritalStatus* contém ambos os valores, de sequência e numérico. Menos de cinco por cento dos valores são sequências. Com a configuração padrão de 95% para **Porcentagem de valores que determinam o formato de dados automático**, a variável é tratada como numérica e os valores de sequência são configurados como omissos do sistema. Se nenhum formato de dados atender ao valor de porcentagem, a variável será tratada como uma variável de sequência. Se você alterar a configuração para 100, todos os valores serão preservados, mas todos os valores numéricos serão tratados como sequências.

15. Clique em **Avançar** para continuar.
16. Deixe as seleções padrão na última etapa e clique em **Concluir** para importar os dados.



## Capítulo 3. Usando o Editor de Dados

O Editor de Dados exibe o conteúdo do arquivo de dados ativos. As informações no Editor de Dados consistem em variáveis e casos.

- Na Visualização de Dados, as colunas representam variáveis e as linhas representam casos (observações).
- Na Visualização de Variável, cada linha é uma variável e cada coluna é um atributo que é associado a essa variável.

As variáveis são usadas para representar os diferentes tipos de dados que você compilou. Uma analogia comum é a de uma pesquisa. A resposta a cada pergunta em uma pesquisa é equivalente a uma variável. As variáveis vêm em muitos tipos diferentes, incluindo números, sequências, moeda e datas.

### Inserindo dados numéricos

Os dados podem ser inseridos no Editor de Dados, o que pode ser útil para arquivos de dados pequenos ou para fazer edições menores em arquivos de dados maiores.

1. Clique na guia **Visualização de Variável** na parte inferior da janela Editor de Dados.

É necessário definir as variáveis que serão usadas. Neste caso, apenas três variáveis são necessárias: *idade*, *estado civil* e *receita*.

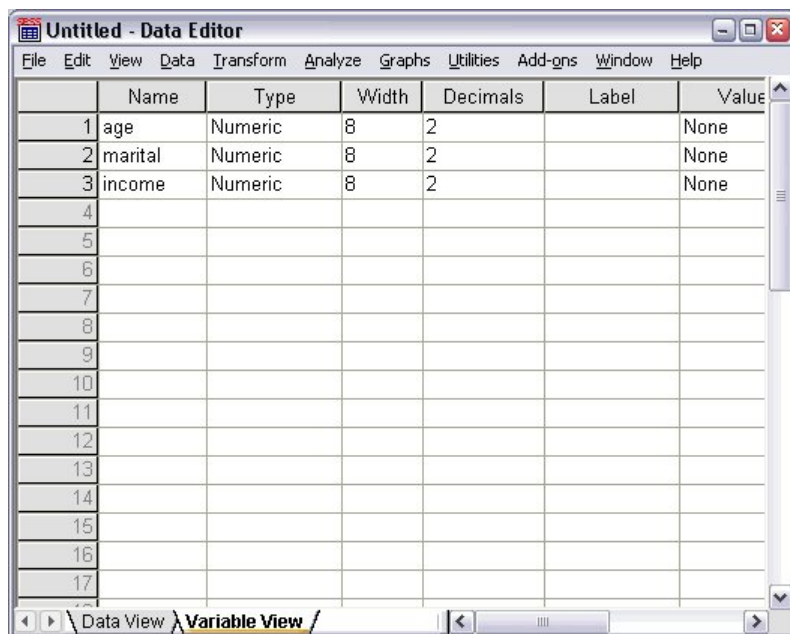


Figura 19. Nomes de variáveis na Visualização de Variável

2. Na primeira linha da primeira coluna, digite *idade*.
3. Na segunda linha, digite *estado civil*.
4. Na terceira linha, digite *receita*.

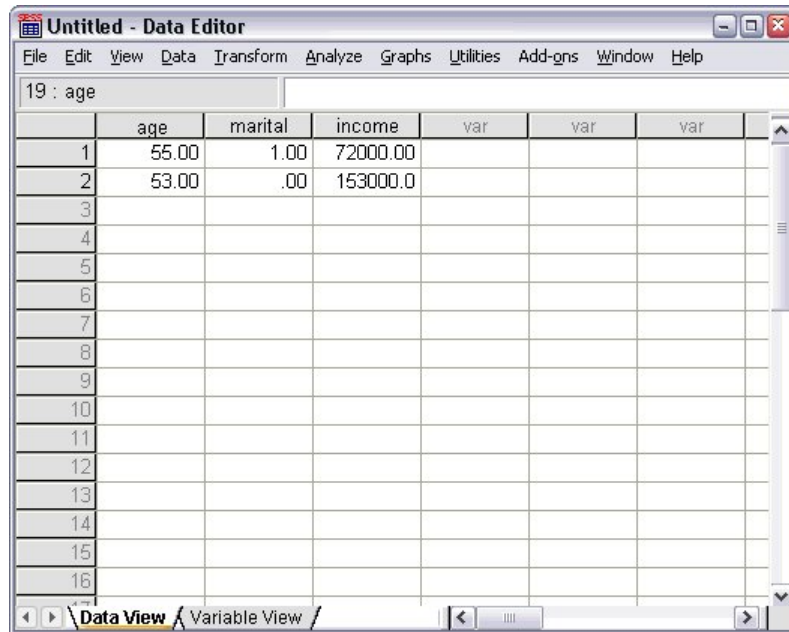
Novas variáveis recebem automaticamente um tipo de dados numéricos.

Se você não inserir nome de variáveis, nomes exclusivos serão automaticamente criados. No entanto, esses nomes não são descritivos e não são recomendados para arquivos de dados grandes.

5. Clique na guia **Visualização de Dados** para continuar inserindo os dados.

Os nomes que você inseriu na Visualização de Variável são, agora, os títulos para as três primeiras colunas na Visualização de Dados.

Comece inserindo dados na primeira linha, iniciando na primeira coluna.



The screenshot shows the 'Untitled - Data Editor' window in SPSS. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The title bar indicates '19 : age'. The data grid has columns labeled 'age', 'marital', 'income', and three 'var' columns. The first two rows contain data: Row 1 has 55.00 for age, 1.00 for marital, and 72000.00 for income. Row 2 has 53.00 for age, .00 for marital, and 153000.0 for income. The bottom status bar shows 'Data View' and 'Variable View' tabs.

	age	marital	income	var	var	var
1	55.00	1.00	72000.00			
2	53.00	.00	153000.0			
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						

Figura 20. Valores inseridos na Visualização de Dados

6. Na coluna *idade*, digite 55.
7. Na coluna *estado civil*, digite 1.
8. Na coluna *receita*, digite 72000.
9. Mova o cursor para a segunda linha da primeira coluna para incluir os dados da próxima pessoa.
10. Na coluna *idade*, digite 53.
11. Na coluna *estado civil*, digite 0.
12. Na coluna *receita*, digite 153000.

Atualmente, as colunas *idade* e *estado civil* exibem pontos decimais, embora seus valores sejam planejados para serem números inteiros. Para ocultar os pontos decimais nessas variáveis:

13. Clique na guia **Visualização de Variável** na parte inferior da janela Editor de Dados.
14. Na coluna *Decimais* da linha *idade*, digite 0 para ocultar o decimal.
15. Na coluna *Decimais* da linha *estado civil*, digite 0 para ocultar o decimal.

## Inserindo dados de sequência

Dados não numéricos, como sequências de texto, também podem ser inseridos no Editor de Dados.

1. Clique na guia **Visualização de Variável** na parte inferior da janela Editor de Dados.
2. Na primeira célula da primeira linha vazia, digite *sexo* para o nome de variável.
3. Clique na célula *Tipo* próxima à entrada.
4. Clique no botão no lado direito da célula *Tipo* para abrir a caixa de diálogo Tipo de Variável.
5. Selecione **Sequência** para especificar o tipo de variável.
6. Clique em **OK** para salvar sua seleção e retornar ao Editor de Dados.



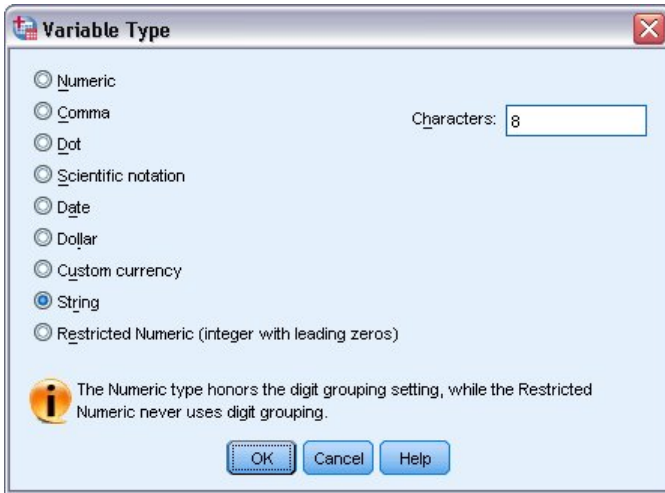


Figura 21. Caixa de diálogo Tipo de Variável

## Definindo dados

Além de definir os tipos de dados, você também pode definir rótulos de variáveis descritivos e rótulos de valor para os nomes de variáveis e os valores de dados. Esses rótulos descritivos são usados em gráficos e relatórios estatísticos.

### Incluindo rótulos de variáveis

Os rótulos são destinados a fornecer descrições de variáveis. Essas descrições são, frequentemente, versões mais longas de nomes de variáveis. Os rótulos podem ter até 255 bytes. Esses rótulos são usados em sua saída para identificar as variáveis diferentes.

1. Clique na guia **Visualização de Variável** na parte inferior da janela Editor de Dados.
2. Na coluna *Rótulo* da linha *idade*, digite *Idade do Entrevistado*.
3. Na coluna *Rótulo* da linha *estado civil*, digite *Estado Civil*.
4. Na coluna *Rótulo* da linha *receita*, digite *Renda Familiar*.
5. Na coluna *Rótulo* da linha *sexo*, digite *Sexo*.

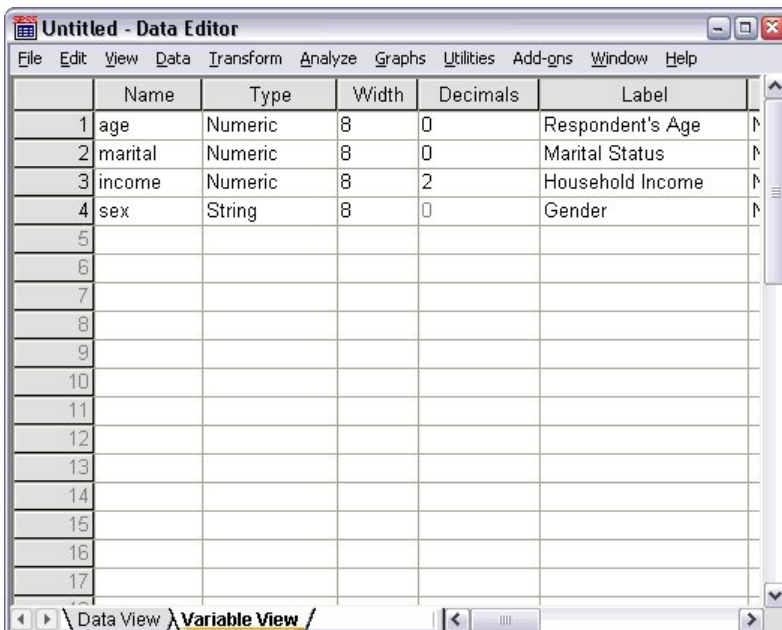


Figura 22. Rótulos de variáveis inseridos na Visualização de Variável

## Alterando o formato e o tipo de variável

A coluna *Tipo* exibe o tipo de dados atuais para cada variável. Os tipos de dados mais comuns são numéricos e de sequência, mas muitos outros formatos são suportados. No arquivo de dados atuais, a variável *receita* é definida como um tipo numérico.

1. Clique na célula *Tipo* para a linha *receita* e, em seguida, clique no botão no lado direito da célula para abrir a caixa de diálogo Tipo de Variável.
2. Selecione **Dólar**.

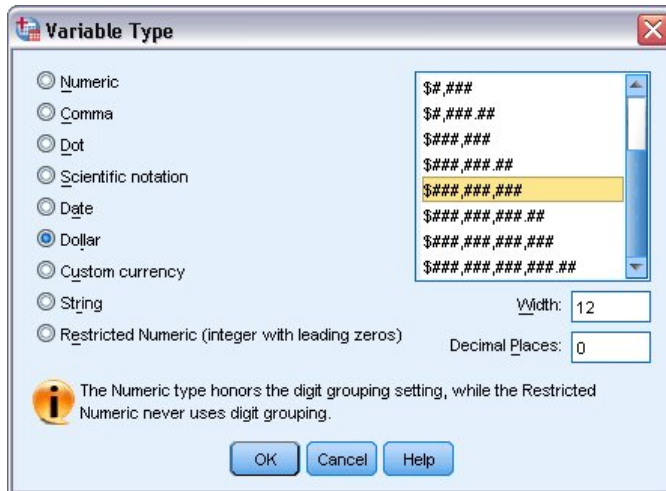


Figura 23. Caixa de diálogo Tipo de Variável

As opções de formatação para o tipo de dados selecionado atualmente são exibidas.

3. Para o formato da moeda neste exemplo, selecione **\$###,###,###**.
4. Clique em **OK** para salvar suas mudanças.

## Incluindo rótulos de valor

Os rótulos de valor fornecem um método para o mapeamento dos valores da variável para um rótulo da sequência. Neste exemplo, há dois valores aceitáveis para a variável *marital*. Um valor de 0 significa que a pessoa é solteira e um valor de 1 significa que ele(a) é casado(a).

1. Clique na célula *Valores* para a linha *marital* e, em seguida, clique no botão no lado direito da célula para abrir a caixa de diálogo Rótulos de Valor.  
O **valor** é o valor numérico real.  
O **rótulo de valor** é o rótulo de sequência que é aplicado ao valor numérico especificado.
2. Digite 0 no campo Valor.
3. Digite Solteiro no campo Rótulo.
4. Clique em **Incluir** para incluir este rótulo na lista.

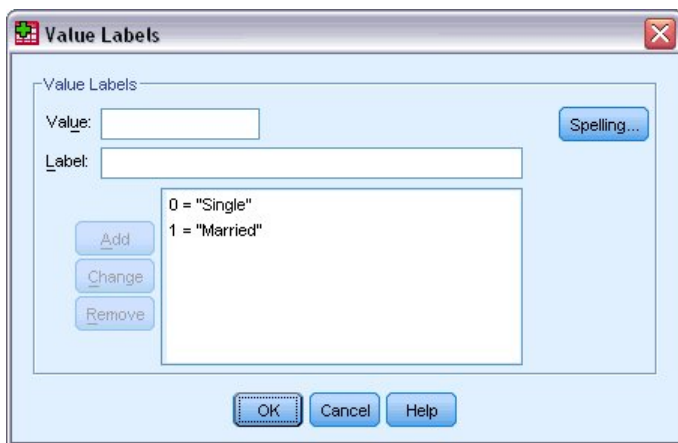


Figura 24. Caixa de diálogo Rótulos de Valor

5. Digite 1 no campo Valor e digite Casado no campo Rótulo.
6. Clique em **Incluir** e, em seguida, clique em **OK** para salvar suas mudanças e retornar ao Editor de Dados.

Esses rótulos podem também ser exibidos na Visualização de Dados, que pode tornar seus dados mais legíveis.

7. Clique na guia **Visualização de Dados** na parte inferior da janela Editor de Dados.
8. Nos menus, escolha:

#### **Visualizar > Rótulos de valor**

Agora, os rótulos são exibidos em uma lista quando você insere valores no Editor de Dados. Esta configuração tem o benefício de sugerir uma resposta válida e fornecer uma resposta mais descritiva.

Se o item de menu Rótulos de Valor já estiver ativo (com um visto junto a ele), escolher **Rótulos de Valor** irá *desligar* a exibição dos rótulos de valor.

## **Manipulando dados omissos**

Dados omissos ou inválidos são, geralmente, muito comuns para ignorar. Os entrevistados podem se recusar a responder algumas perguntas, podem não saber responder ou podem responder em um formato inesperado. Se você não filtrar ou identificar esses dados, sua análise pode não fornecer resultados precisos.

Para dados numéricos, os campos de dados vazios ou os campos que contêm entradas inválidas são convertidos em omissos do sistema, que são identificáveis por um único período.

A razão pela qual um valor está omissos pode ser importante para sua análise. Por exemplo, você pode achar útil distinguir entre os entrevistados que se recusaram a responder a uma pergunta e os entrevistados que não responderam a uma pergunta porque ela não era aplicável.

## **Valores omissos para uma variável numérica**

1. Clique na guia **Visualização de Variável** na parte inferior da janela Editor de Dados.
2. Clique na célula *Omisso* na linha *idade* e, em seguida, clique no botão no lado direito da célula para abrir a caixa de diálogo Valores Omissos.

Nesta caixa de diálogo, é possível especificar até três valores omissos distintos ou especificar um intervalo de valores mais um valor discreto adicional.



Figura 25. Caixa de diálogo Valores Omissos

3. Selecione **Valores omissos discretos**.
4. Digite 999 na primeira caixa de texto e deixe as outras duas caixas de texto vazias.
5. Clique em **OK** para salvar suas mudanças e retornar ao Editor de Dados.  
Agora que o valor de dados omissos foi incluído, um rótulo pode ser aplicado a esse valor.
6. Clique na célula *Valores* na linha *idade* e, em seguida, clique no botão no lado direito da célula para abrir a caixa de diálogo Rótulos de Valor.
7. Digite 999 no campo Valor.
8. Digite Nenhuma Resposta no campo Rótulo.
9. Clique em **Incluir** para incluir esse rótulo no arquivo de dados.
10. Clique em **OK** para salvar suas mudanças e retornar ao Editor de Dados.

## Valores omissos para uma variável de sequência

Os valores omissos para variáveis de sequência são manipulados de forma semelhante aos valores omissos para variáveis numéricas. No entanto, diferente das variáveis numéricas, os campos vazios nas variáveis de sequência não são designados como omissos do sistema. Em vez disso, eles são interpretados como uma sequência vazia.

1. Clique na guia **Visualização de Variável** na parte inferior da janela Editor de Dados.
2. Clique na célula *Omissos* na linha *sexo* e, em seguida, clique no botão no lado direito da célula para abrir a caixa de diálogo Valores Omissos.
3. Selecione **Valores omissos discretos**.
4. Digite NR na primeira caixa de texto.  
Os valores omissos para variáveis de sequência fazem distinção entre maiúsculas e minúsculas. Portanto, um valor de *nr* não é tratado como um valor omissos.
5. Clique em **OK** para salvar suas mudanças e retornar ao Editor de Dados.  
Agora, você pode incluir um rótulo para o valor omissos.
6. Clique na célula *Valores* na linha *sexo* e, em seguida, clique no botão no lado direito da célula para abrir a caixa de diálogo Rótulos de Valor.
7. Digite NR no campo Valor.
8. Digite Nenhuma Resposta no campo Rótulo.
9. Clique em **Incluir** para incluir este rótulo no seu projeto.
10. Clique em **OK** para salvar suas mudanças e retornar ao Editor de Dados.

---

# Capítulo 4. Examinando estatísticas de resumo para variáveis individuais

Esta seção discute medidas simples de resumo e como o nível de medição de uma variável influencia os tipos de estatísticas que devem ser usadas. Vamos usar o arquivo de dados *demo.sav*. Consulte o tópico Capítulo 10, “Arquivos de Amostra”, na página 79 para obter mais informações

---

## Nível de medição

Medidas de resumo diferentes são apropriadas para tipos diferentes de dados, dependendo do nível de medição:

**Catégoricos.** Dados com um número limitado de valores ou categorias distintas (por exemplo, sexo ou estado civil). Também referidos como **dados qualitativos**. As variáveis catégoricas podem ser dados de sequência (alfanuméricos) ou variáveis numéricas que usam códigos numéricos para representar categorias (por exemplo, 0 = *Solteiro* e 1 = *Casado*). Há dois tipos básicos de dados catégoricos:

- **Nominal.** Dados catégoricos em que não há nenhuma ordem inerente para as categorias. Por exemplo, uma categoria de tarefa de *vendas* não é superior ou inferior a uma categoria de tarefa de *marketing* ou *pesquisa*.
- **Ordinal.** Dados catégoricos em que há uma ordem significativa de categorias, mas não há uma distância considerável entre as categorias. Por exemplo, há uma ordem para os valores *alto*, *médio* e *baixo*, mas a "distância" entre os valores não pode ser calculada.

**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 referido como **quantitativo** ou **dados contínuos**.

---

## Medidas de resumo para dados catégoricos

Para dados catégoricos, a medida de resumo mais típica é o número ou a porcentagem de casos em cada categoria. O **modo** é a categoria com o maior número de casos. Para dados ordinais, o **mediano** (o valor no qual metade dos casos se encaixa acima ou abaixo) poderá também ser uma medida de resumo útil se houver um grande número de categorias.

O procedimento Frequências produz tabelas de frequência que exibem o número e a porcentagem de casos para cada valor observado de uma variável.

1. Nos menus, escolha:

**Analisar > Estatística Descritiva > Frequências...**

Nota: Este recurso requer a opção Estatísticas Base.

2. Selecione as opções *Possui PDA [ownpda]* e *Possui TV [owntv]* e mova-as para a lista Variável(is).

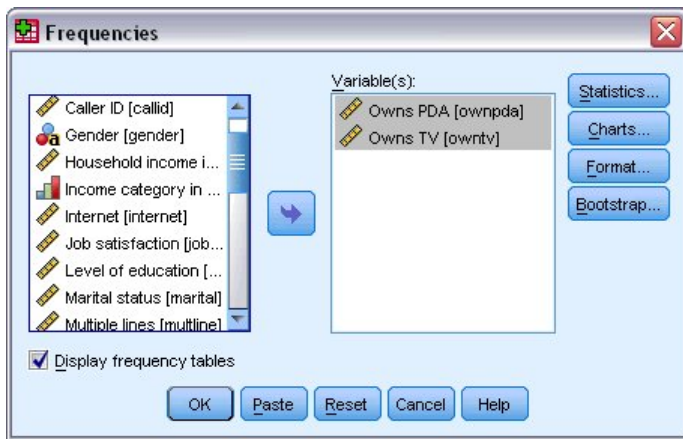


Figura 26. Variáveis categóricas selecionadas para análise

3. Clique em **OK** para executar o procedimento.

**Frequency Table**

**Owns PDA**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	5093	79.6	79.6	79.6
	Yes	1307	20.4	20.4	100.0
Total		6400	100.0	100.0	

**Owns TV**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	63	1.0	1.0	1.0
	Yes	6337	99.0	99.0	100.0
Total		6400	100.0	100.0	

Figura 27. Tabelas de frequência

As tabelas de frequência são exibidas na janela do Visualizador. As tabelas de frequência revelam que apenas 20,4% das pessoas possuem PDAs, mas quase todas possuem uma TV (99,0%). Estas podem não ser revelações interessantes, embora possa ser interessante descobrir mais sobre o pequeno grupo de pessoas que não possuem televisões.

## Gráficos para dados categóricos

É possível exibir graficamente as informações em uma tabela de frequências com um gráfico de barras ou um gráfico de pizza.

1. Abra a caixa de diálogo Frequências novamente. (As duas variáveis ainda devem ser selecionadas.)

Você pode usar o botão Reclamada de Diálogo na barra de ferramentas para retornar rapidamente aos procedimentos usados recentemente.



Figura 28. Botão Reclamada de Diálogo

2. Clique em **Gráficos**.
3. Selecione **Gráficos de barras** e, em seguida, clique em **Continuar**.
4. Clique em **OK** na caixa de diálogo principal para executar o procedimento.

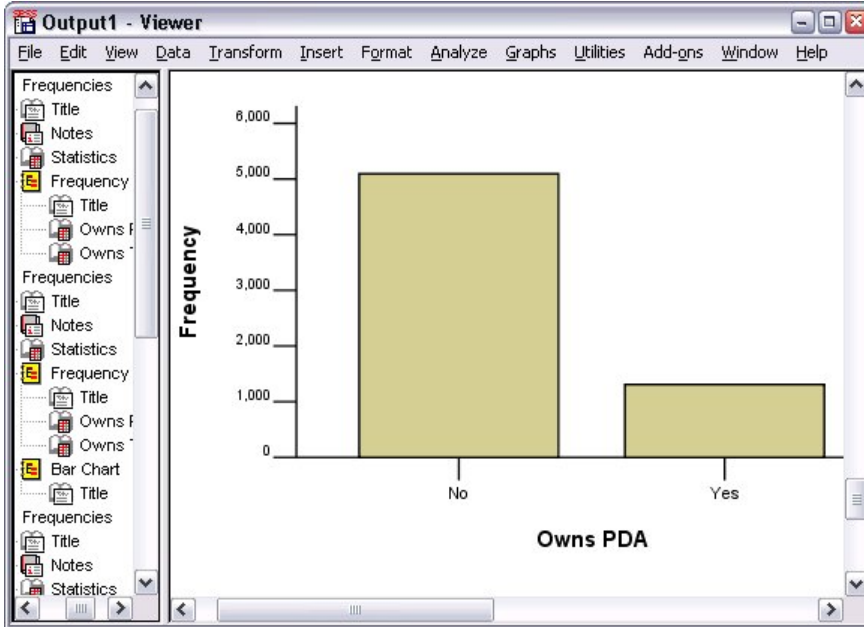


Figura 29. Gráfico de barras

Além das tabelas de frequências, as mesmas informações agora são exibidas na forma de gráficos de barras, tornando fácil ver que a maioria das pessoas não possui PDAs, mas quase todas possuem uma TV.

## Medidas de resumo para variáveis de escala

Há muitas medidas de resumo disponíveis para variáveis de escala, incluindo:

- **Medidas de tendência central.** As medidas mais comuns de tendência central são a **média** (média aritmética) e a **mediana** (valor no qual metade dos casos se encaixam acima e abaixo).
- **Medidas de dispersão.** Estatísticas que medem a quantidade de variação ou difusão nos dados incluem o desvio padrão, mínimo e máximo.

1. Abra a caixa de diálogo Frequências novamente.
2. Clique em **Reconfigurar** para limpar as configurações anteriores.
3. Selecione *Receita familiar em milhares [receita]* e mova esta opção para a lista Variável(is).
4. Clique em **Estatísticas**.
5. Selecione **Média, Mediana, Desvio padrão, Mínimo e Máximo**.
6. Clique em **Continuar**.
7. Cancele a seleção de **Exibir tabelas de frequências** na caixa de diálogo principal. (As tabelas de frequências geralmente não são úteis para variáveis de escala, uma vez que pode haver quase tantos valores distintos quanto há casos no arquivo de dados.)
8. Clique em **OK** para executar o procedimento.

A tabela Estatísticas de Frequências é exibida na janela Visualizador.

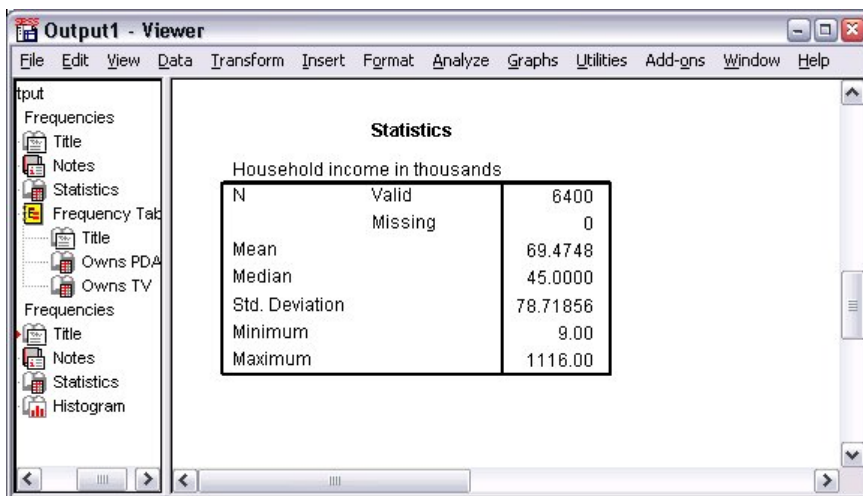


Figura 30. Tabela Estatísticas de Frequências

Neste exemplo, há uma grande diferença entre o médio e o mediano. O médio é quase 25.000 maior que o mediano, indicando que os valores não são normalmente distribuídos. É possível verificar visualmente a distribuição com um histograma.

## Histogramas para variáveis de escala

1. Abra a caixa de diálogo Frequências novamente.
2. Clique em **Gráficos**.
3. Selecione **Histogramas** e **Com curva normal**.
4. Clique em **Continuar** e, em seguida, clique em **OK** na caixa de diálogo principal para executar o procedimento.

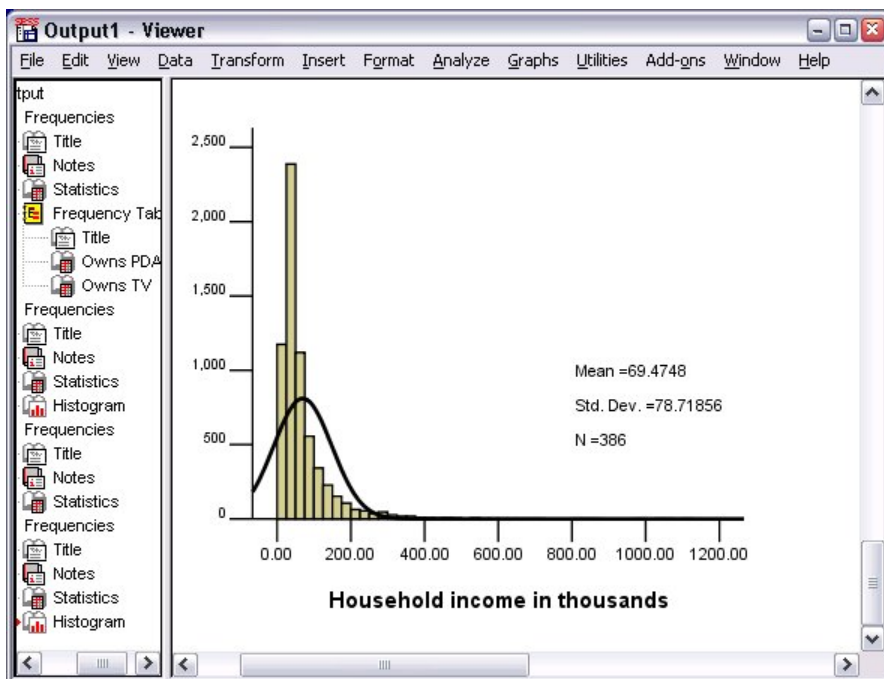


Figura 31. Histograma

A maioria dos casos é armazenada em cluster na parte final inferior da escala, com a maioria ficando abaixo de 100.000. Há, no entanto, alguns casos no intervalo de 500.000 e além (muito poucos para serem visíveis sem modificar o histograma). Esses valores altos para apenas alguns casos têm um efeito



significativo na média, mas pouco ou nenhum efeito na mediana, tornando a mediana um indicador melhor de tendência central neste exemplo.



# Capítulo 5. Criando e editando gráficos

Você pode criar e editar uma ampla variedade de tipos de gráficos. Neste capítulo, vamos criar e editar gráficos de barras. É possível aplicar os princípios a qualquer tipo de gráfico.

## Fundamentos para criação de gráfico

Para demonstrar os fundamentos da criação de gráfico, vamos criar um gráfico de barras de receita média para níveis diferentes de satisfação da tarefa. Este exemplo usa o arquivo de dados *demo.sav*. Consulte o tópico Capítulo 10, “Arquivos de Amostra”, na página 79 para obter mais informações

1. Nos menus, escolha:

### Gráficos > Construtor de Gráfico...

A caixa de diálogo Construtor de Gráfico é uma janela interativa que permite visualizar como será a aparência de um gráfico enquanto você o constrói.

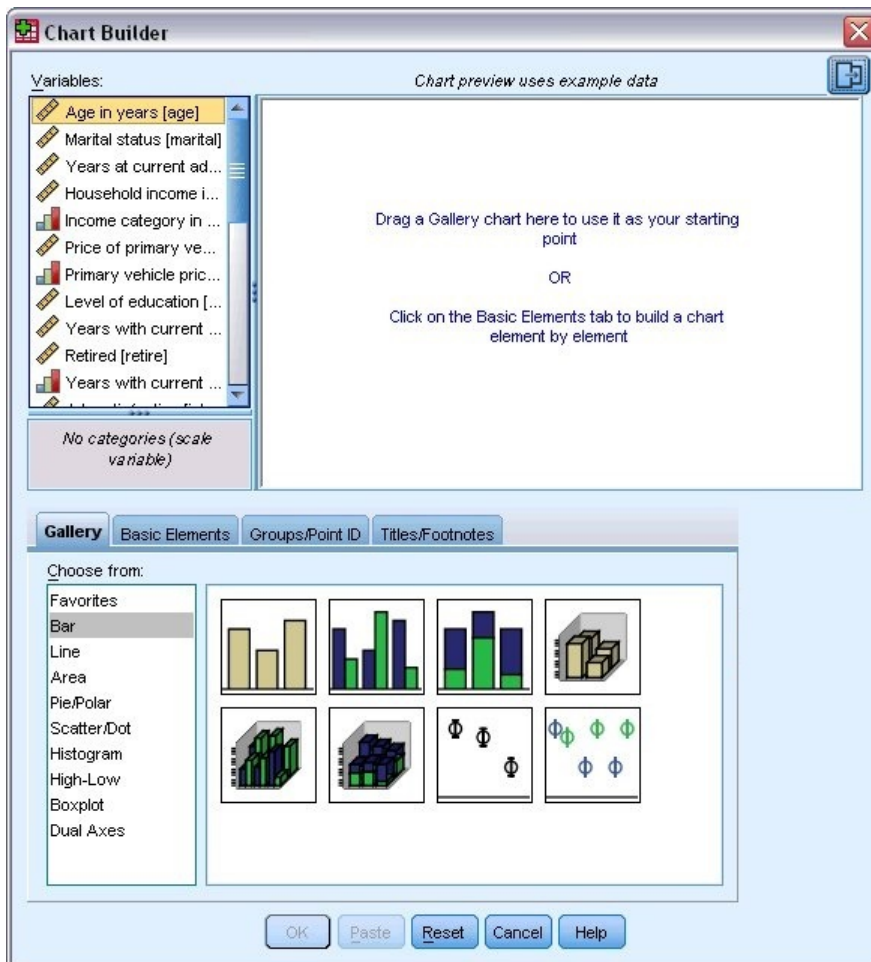


Figura 32. Caixa de diálogo Construtor de Gráfico

## Usando a galeria do Construtor de Gráfico

1. Clique na guia **Galeria** se ela não estiver selecionada.

A Galeria inclui muitos gráficos predefinidos diferentes, que são organizados por tipo de gráfico. A guia Elementos Básicos também fornece elementos básicos (como eixos e elementos gráficos) para criar gráficos a partir do zero, mas é mais fácil usar a Galeria.

2. Clique em **Barra** se esta opção não estiver selecionada.

Ícones que representam os gráficos de barras disponíveis na Galeria aparecem na caixa de diálogo. As figuras devem fornecer informações suficientes para identificar o tipo de gráfico específico. Se você precisar de mais informações, poderá também exibir a descrição Dica de Ferramenta do gráfico mantendo o cursor sobre um ícone.

3. Arraste o ícone para o gráfico de barras simples na "tela", que é a área grande acima da Galeria. O Construtor de Gráfico exibe uma visualização do gráfico na tela. Observe que os dados usados para traçar o gráfico não são seus dados reais. Eles são dados de exemplo.

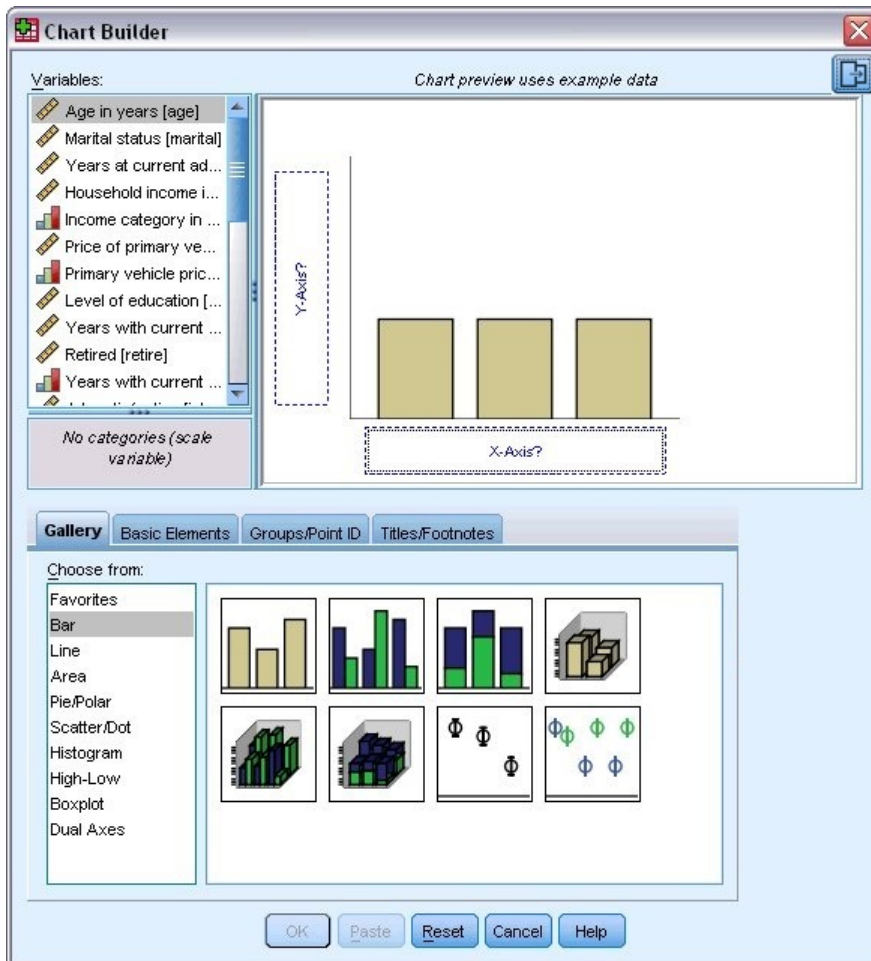


Figura 33. Gráfico de barras na tela do Construtor de Gráfico

## Definindo variáveis e estatísticas

Embora haja um gráfico na tela, ele não é completo porque não há variáveis ou estatísticas para controlar a altura das barras e para especificar qual categoria de variável corresponde a cada barra. Você não pode ter um gráfico sem variáveis e estatísticas. Você inclui variáveis arrastando-as da lista Variáveis, que está localizada à esquerda da tela.

O nível de medição de uma variável é importante no Construtor de Gráfico. Você vai usar a variável *Satisfação de tarefa* no eixo x. No entanto, o ícone (que é semelhante a uma régua) próximo à variável indica que seu nível de medição é definido como escala. Para criar o gráfico correto, você deve usar um nível de medição categórico. Em vez de voltar e alterar o nível de medição na Visualização de Variável, você pode alterar o nível de medição temporariamente no Construtor de Gráfico.

1. Clique com o botão direito *Satisfação de tarefa* na lista Variáveis e escolha **Ordinal**. Ordinal é um nível de medição apropriado porque as categorias em *Satisfação de tarefa* podem ser classificadas por nível de satisfação. Observe que o ícone muda depois de você mudar o nível de medição.
2. Agora, arraste *Satisfação de tarefa* da lista Variáveis para a área de lançamento do eixo x.  
A área de lançamento do eixo y é padronizada para a estatística *Contagem*. Se desejar usar outra estatística (como porcentagem ou média), você poderá facilmente alterá-la. Você não usará nenhuma dessas estatísticas neste exemplo, mas nós vamos revisar o processo no caso de você precisar alterar esta estatística em outro momento.
3. Clique na guia **Propriedades do elemento** na barra lateral do Construtor de gráfico. (Se a barra lateral não for exibida, clique no botão no canto superior direito do Construtor de gráfico para exibir a barra lateral.)

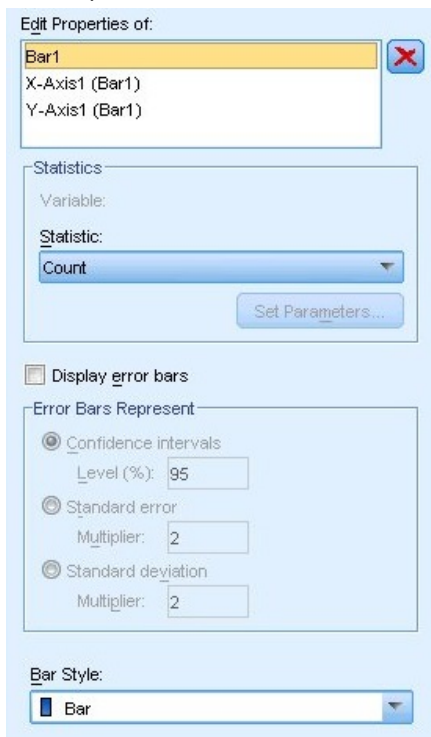


Figura 34. Propriedades do elemento

Propriedades do elemento permite alterar as propriedades de vários elementos do gráfico. Esses elementos incluem os elementos de gráficos (como as barras no gráfico de barras) e os eixos no gráfico. Selecione um dos elementos em Editar Propriedades da lista para alterar as propriedades associadas a esse elemento. Observe também o X vermelho localizado à direita da lista. Este botão exclui um elemento gráfico da tela. Como **Bar1** é selecionado, as propriedades mostradas aplicam-se aos elementos gráficos, especificamente ao elemento gráfico de barra.

A lista suspensa Estatística mostra as estatísticas específicas que estão disponíveis. As mesmas estatísticas estão, geralmente, disponíveis para cada tipo de gráfico. Esteja ciente de que algumas estatísticas requerem que a área de lançamento do eixo y contenha uma variável.

4. Arraste *Receita doméstica em milhares* da lista Variáveis para a área de lançamento do eixo y. Como a variável no eixo y é escalar e a variável do eixo x é categórica (ordinal é um tipo de nível de medição categórica), a área de lançamento do eixo y é padronizada para a estatística *Média*. Essas são as variáveis e estatísticas que você deseja, portanto, não é necessário alterar as propriedades do elemento.

## Incluindo Texto

Você também pode incluir títulos e notas de rodapé no gráfico.

1. Clique na guia **Títulos/Notas de Rodapé**.

## 2. Selecione **Título 1**.

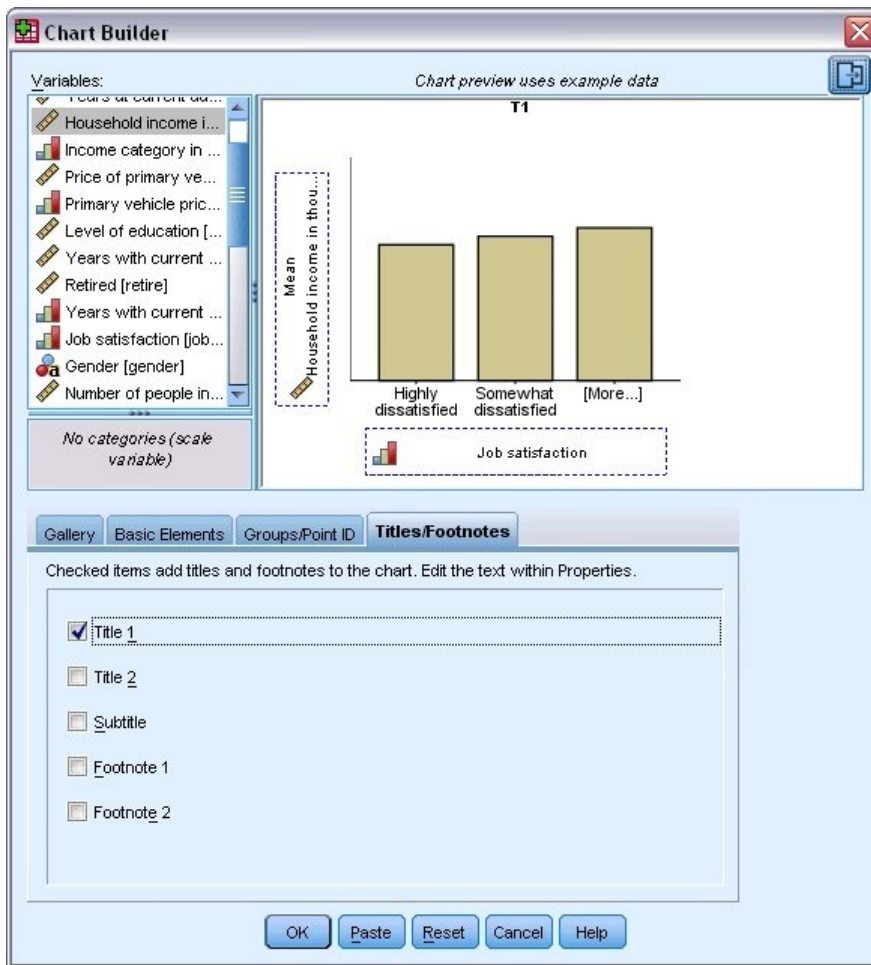


Figura 35. Título 1 exibido na tela

O título aparece na tela com o rótulo **T1**.

3. Na guia **Propriedades do elemento**, selecione **Título 1** em Editar Propriedades da lista.
4. Na caixa de texto Conteúdo, digite Receita por Satisfação da Tarefa. Este é o texto que o título exibirá.

## Criando o gráfico

1. Clique em **OK** para criar o gráfico de barras.

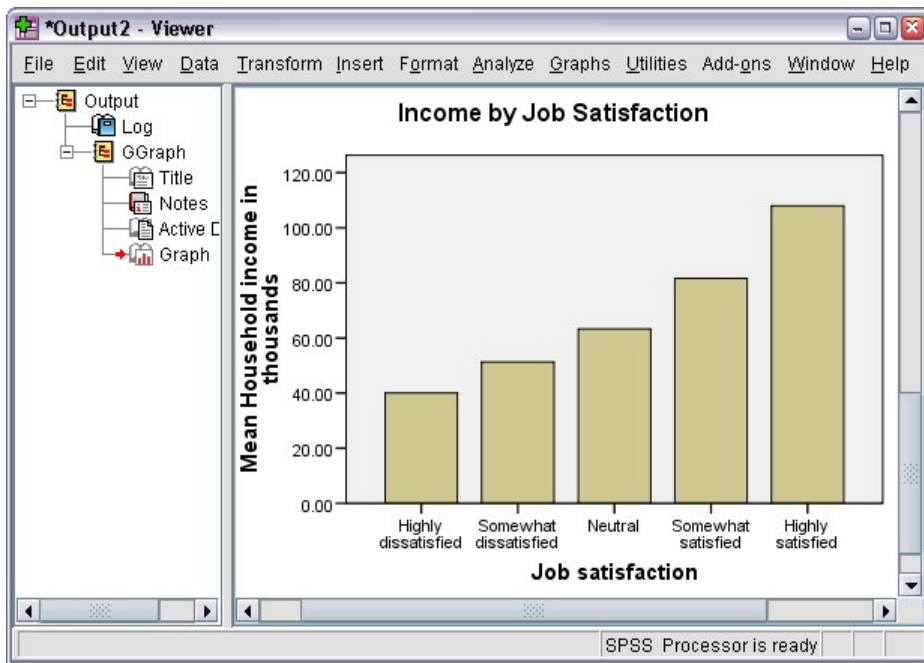


Figura 36. Gráfico de barras

O gráfico de barras revela que os respondentes que estão mais satisfeitos com seus trabalhos tendem a ter receitas domésticas mais altas.





## Capítulo 6. Trabalhando com a Saída

Os resultados da execução de um procedimento estatístico são exibidos no Visualizador. A saída produzida pode ser tabelas estatísticas, diagramas, gráficos ou texto, dependendo das opções feitas quando você executa o procedimento. Esta seção usa os arquivos *viewertut.spv* e *demo.sav*. Consulte o tópico [Capítulo 10, “Arquivos de Amostra”, na página 79](#) para obter mais informações

### Usando o Visualizador

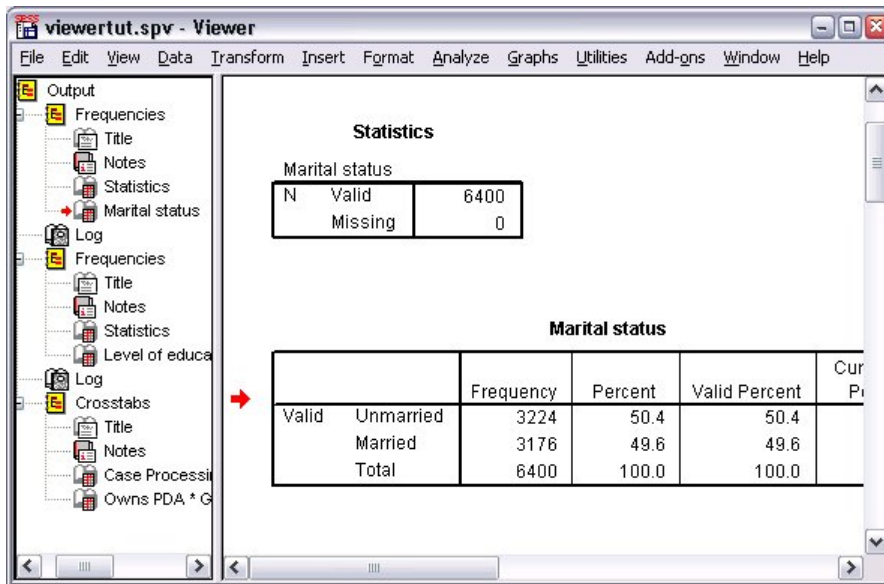


Figura 37. Visualizador

A janela do Visualizador é dividida em duas áreas de janela. A **área de janela da estrutura de tópicos** contém uma estrutura de tópicos de todas as informações armazenadas no Visualizador. A **área de janela de conteúdo** contém tabelas estatísticas, gráficos e saída de texto.

Use as barras de rolagem para navegar pelo conteúdo da janela, vertical e horizontalmente. Para uma navegação mais fácil, clique em um item na área de janela da estrutura de tópicos para exibi-lo na área de janela de conteúdo.

1. Clique e arraste a borda direita da área de janela da estrutura de tópicos para mudar sua largura.

Um ícone de livro aberto na área de janela da estrutura de tópicos indica que ele está atualmente visível no Visualizador, embora ele não possa estar atualmente na parte visível da área de janela de conteúdo.

2. Para ocultar uma tabela ou gráfico, clique duas vezes em seu ícone de livro na área de janela da estrutura de tópicos.

O ícone de livro aberto muda para um ícone de livro fechado, o que significa que as informações associadas a ele agora estão ocultas.

3. Para exibir novamente a saída oculta, clique duas vezes no ícone de livro fechado.

É possível também ocultar tudo da saída a partir de um procedimento estatístico particular ou tudo da saída no Visualizador.

4. Clique na caixa com o sinal de menos (-) à esquerda do procedimento cujos resultados você deseja ocultar ou clique na caixa próxima do item mais acima na área de janela da estrutura de tópicos para ocultar toda a saída.

A estrutura de tópicos é reduzida, indicando visualmente que esses resultados estão ocultos.

É possível também mudar a ordem na qual a saída é exibida.

5. Na área de janela da estrutura de tópicos, clique nos itens que deseja mover.
6. Arraste os itens selecionados para um novo local na estrutura de tópicos.

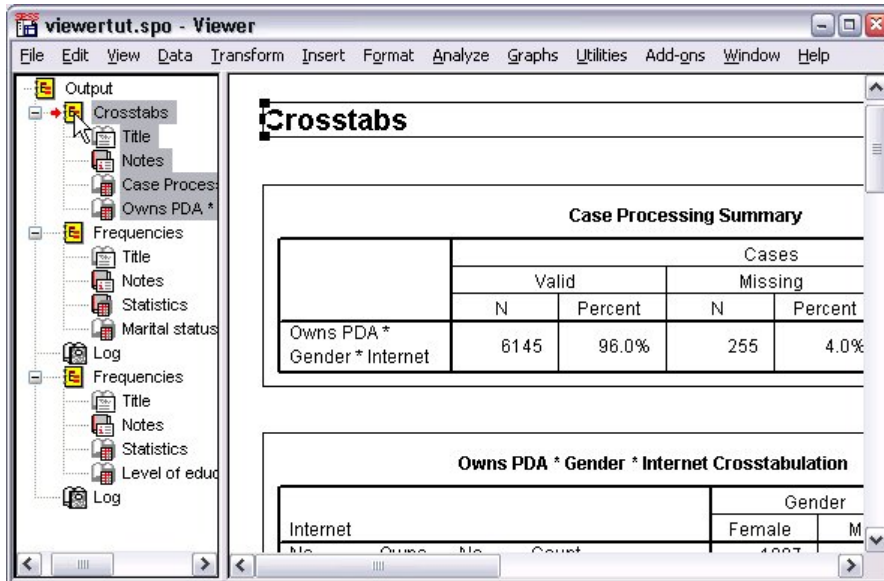


Figura 38. Saída reordenada no Visualizador

É possível também mover itens de saída clicando neles e arrastando-os na área de janela de conteúdo.

## Usando o Editor de Tabela Dinâmica

Os resultados da maioria dos procedimentos estatísticos são exibidos nas **tabelas dinâmicas**.

### Acessando definições de saída

Muitos termos estatísticos são exibidos na saída. As definições desses termos podem ser acessadas diretamente no Visualizador.

1. Clique duas vezes na tabela *Possui PDA \* Sexo \* Tabulação Cruzada da Internet*.
2. Clique com o botão direito em *Contagem Esperada* e escolha **O que é isso?** no menu pop-up.

A definição é exibida em uma janela pop-up.

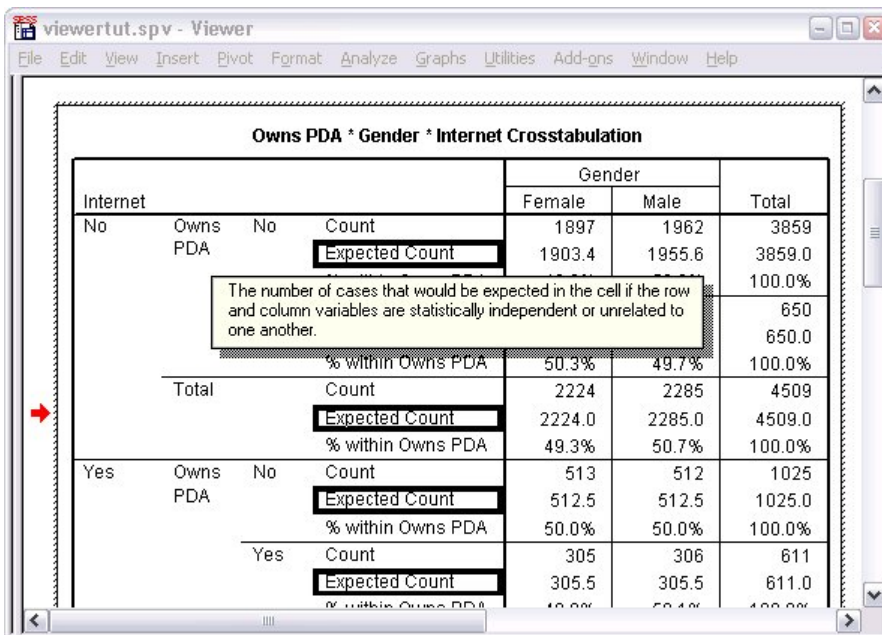


Figura 39. Definição pop-up

## Tabelas dinâmicas

As tabelas padrão produzidas podem não exibir informações tão organizadamente ou tão claramente quanto você gostaria. Com as tabelas dinâmicas, é possível transpor linhas e colunas ("inverter" a tabela), ajustar a ordem de dados em uma tabela e modificar a tabela de muitas outras maneiras. Por exemplo, é possível transformar uma tabela curta, larga em uma tabela longa, fina transpondo linhas e colunas. Mudar o layout da tabela não afeta os resultados. Em vez disso, é uma maneira de exibir suas informações de uma maneira diferente ou mais desejável.

1. Se ela ainda não estiver ativada, clique duas vezes na tabela *Possui PDA \* Sexo \* Tabulação Cruzada da Internet* para ativá-la.
2. Se a janela Bandejas de Dinamização não estiver visível, nos menus, escolha:

### Dinâmica > Bandejas de Dinamização

As bandejas de dinamização fornecem uma maneira de mover dados entre colunas, linhas e estratos.

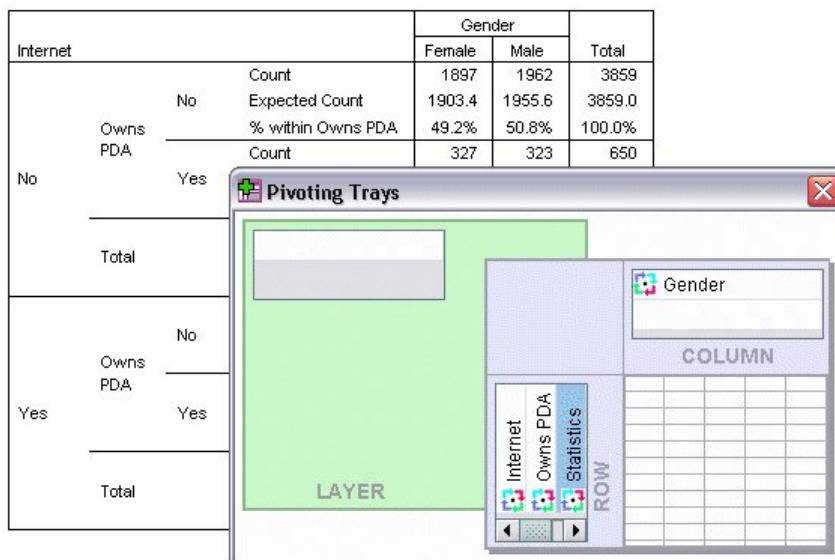


Figura 40. Bandejas de dinamização

3. Arraste o elemento *Estatísticas* da dimensão da Linha para a dimensão da Coluna, abaixo de *Sexo*. A tabela é imediatamente reconfigurada para refletir suas mudanças.

A ordem dos elementos na bandeja de dinamização reflete a ordem dos elementos na tabela.

4. Arraste e solte o elemento *Possui PDA* antes do elemento *Internet* na dimensão da linha para inverter a ordem dessas duas linhas.

			Gender					Total		
			Female			Male				
			Count	Expected Count	% within Owns PDA	Count	Expected Count	% within Owns PDA		
Owns PDA	No	No	1897	1903.4	49.2%				3859.0	100.0%
	Yes	Yes	1025.0						1025.0	100.0%
Internet	No	No	650.0						650.0	100.0%
	Yes	Yes	611.0						611.0	100.0%
Total		No	2447.0						4509.0	100.0%
		Yes	1636.0						1636.0	100.0%

Figura 41. Trocar linhas

## Criando e exibindo estratos

As estratos podem ser úteis para tabelas grandes com categorias aninhadas de informações. Ao criar estratos, você simplifica a aparência da tabela, tornando-a mais fácil de ler.

1. Arraste o elemento *Sexo* da dimensão da Coluna para a dimensão da Camada.

Gender		Female		
		Count	Expected Count	% within Owns PDA
Owns PDA	No	1897	1903.4	49.2%
	Yes			
Internet	No			
	Yes			
Total				

Figura 42. Ícone dinâmico de Sexo na dimensão da Camada

Para exibir uma estrato diferente, selecione uma categoria na lista suspensa na tabela.

## Editando tabelas

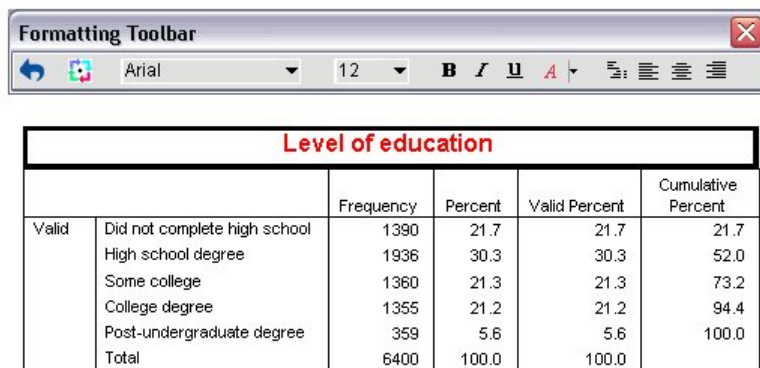
A menos que você tenha reservado o tempo para criar um TableLook customizado, as tabelas dinâmicas são criadas com formatação padrão. É possível mudar a formatação de qualquer texto dentro de uma

tabela. Os formatos que podem ser mudados incluem nome da fonte, tamanho da fonte, estilo da fonte (negrito ou itálico) e cor.

1. Clique duas vezes na tabela *Nível de Educação*.
2. Se a barra de ferramentas Formatação não estiver visível, nos menus, escolha:

**Visualizar > Barra de ferramentas**

3. Clique no texto do título, *Nível de educação*.
4. Na lista suspensa de tamanhos de fonte na barra de ferramentas, escolha **12**.
5. Para mudar a cor do texto do título, clique na ferramenta de cor de texto e escolha uma nova cor.



The image shows a 'Formatting Toolbar' window with options for font face (Arial), size (12), and bold/italic/underline. Below it is a table with a red title 'Level of education'. The table has columns for Frequency, Percent, Valid Percent, and Cumulative Percent.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not complete high school	1390	21.7	21.7	21.7
	High school degree	1936	30.3	30.3	52.0
	Some college	1360	21.3	21.3	73.2
	College degree	1355	21.2	21.2	94.4
	Post-undergraduate degree	359	5.6	5.6	100.0
Total		6400	100.0	100.0	

Figura 43. Texto do título reformatado na tabela dinâmica

É possível também editar o conteúdo de tabelas e rótulos. Por exemplo, é possível mudar o título desta tabela.

6. Clique duas vezes no título.
7. Digite *Nível de Educação* para o novo rótulo.

*Nota:* Se você mudar os valores em uma tabela, os totais e outras estatísticas não serão recalculados.

## Ocultando linhas e colunas

Alguns dos dados exibidos em uma tabela podem não ser úteis ou podem complicar desnecessariamente a tabela. Felizmente, é possível ocultar linhas e colunas inteiras sem perder nenhum dado.

1. Se ele ainda não estiver ativado, clique duas vezes na tabela *Nível de Educação* para ativá-lo.
2. Clique no rótulo da coluna *Percentual Válido* para selecioná-lo.
3. No menu Editar ou no menu pop-up com clique com o botão direito, escolha:

**Selecionar > Células de dados e de rótulo**

4. No menu Visualizar, escolha **Ocultar** ou no menu pop-up com clique com o botão direito, escolha **Ocultar Categoria**.

A coluna agora está oculta, mas não excluída.

**Education Level**

		Frequency	Percent	Cumulative Percent
Valid	Did not complete high school	1390	21.7	21.7
	High school degree	1936	30.3	52.0
	Some college	1360	21.3	73.2
	College degree	1355	21.2	94.4
	Post-undergraduate degree	359	5.6	100.0
Total		6400	100.0	

Figura 44. Coluna *Percentual Válido* oculta na tabela

Para exibir novamente a coluna:

5. Nos menus, escolha:

**Visualizar > Mostrar tudo**

As linhas podem ser ocultas e exibidas da mesma maneira que as colunas.

## Alterando formatos de exibição de dados

Você pode facilmente alterar o formato de exibição de dados em tabelas dinâmicas.

1. Se ele ainda não estiver ativado, clique duas vezes na tabela *Nível de Educação* para ativá-lo.
2. Clique no rótulo da coluna *Percentual* para selecioná-lo.
3. No menu Editar ou no menu pop-up de clique com o botão direito, escolha:

**Selecionar > Células de dados**

4. No menu Formato ou no menu pop-up com clique com o botão direito, escolha **Propriedades da Célula**.
5. Clique na guia **Valor de Formato**.
6. Digite 0 no campo Decimais para ocultar todos os pontos decimais nesta coluna.

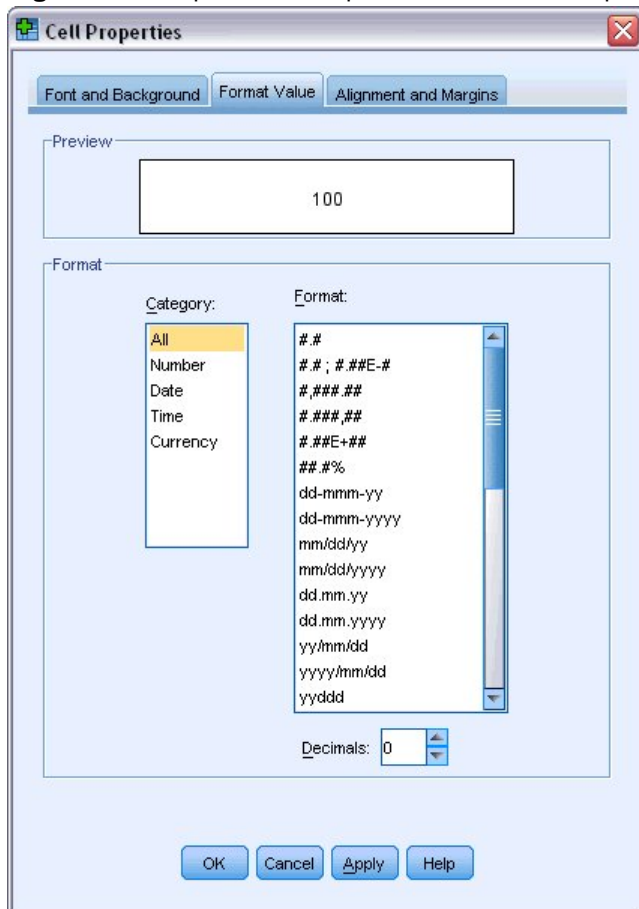


Figura 45. Propriedades da célula, guia Valor de Formato

É possível também alterar o formato e o tipo de dados nesta caixa de diálogo.

7. Selecione o tipo que você deseja na lista Categoria e, em seguida, selecione o formato para esse tipo na lista Formato.
8. Clique em **OK** ou em **Aplicar** para aplicar suas mudanças.

## Education Level

		Frequency	Percent	Cumulative Percent
Valid	Did not complete high school	1390	22	21.7
	High school degree	1936	30	52.0
	Some college	1360	21	73.2
	College degree	1355	21	94.4
	Post-undergraduate degree	359	6	100.0
	Total	6400	100	

Figura 46. Decimais ocultos na coluna Percentual

Os decimais são, agora, ocultos na coluna *Percentual*.

## TableLooks

O formato de suas tabelas é uma parte crítica do fornecimento de resultados claros, concisos e significativos. Se sua tabela for difícil de ler, as informações contidas nessa tabela podem não ser facilmente entendidas.

### Usando formatos predefinidos

1. Clique duas vezes na tabela *Estado Civil*.
2. Nos menus, escolha:

#### Formatar > TableLooks...

A caixa de diálogo TableLooks lista uma variedade de estilos predefinidos. Selecione um estilo na lista para visualizá-lo na janela Amostra à direita.

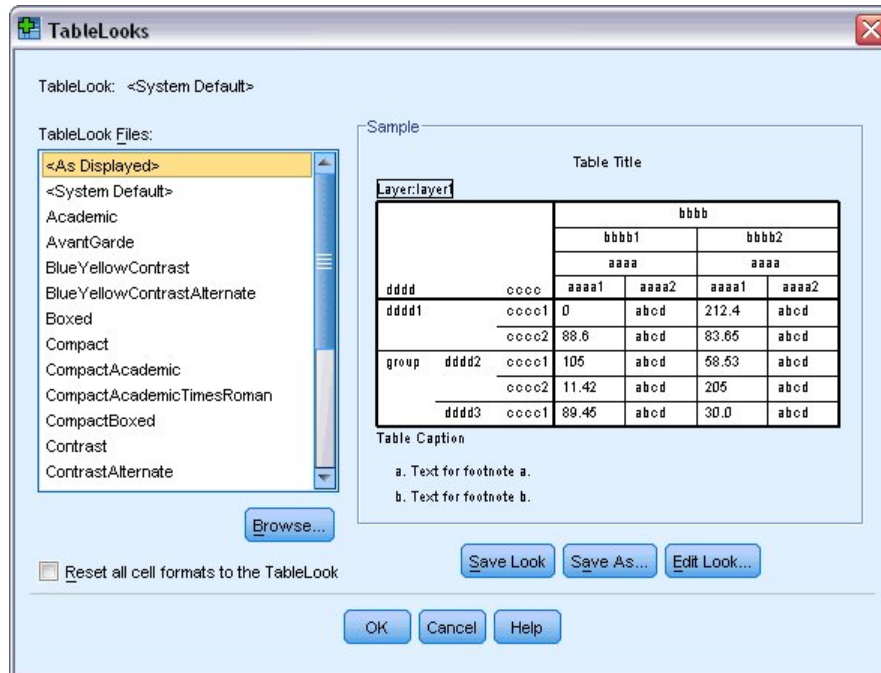


Figura 47. Caixa de diálogo TableLooks

É possível usar um estilo no estado em que se encontra ou editar um estilo existente para atender melhor às suas necessidades.

3. Para usar um estilo existente, selecione um e clique em **OK**.

## Customizando estilos do TableLook

É possível customizar um formato para se ajustar às suas necessidades específicas. Quase todos os aspectos de uma tabela podem ser customizados, da cor do plano de fundo aos estilos de borda.

1. Clique duas vezes na tabela *Estado Civil*.
2. Nos menus, escolha:

**Formatar > TableLooks...**

3. Selecione o estilo que está mais próximo ao formato que você deseja e clique em **Editar Aparência**.
4. Clique na guia **Formatos de Célula** para visualizar as opções de formatação.

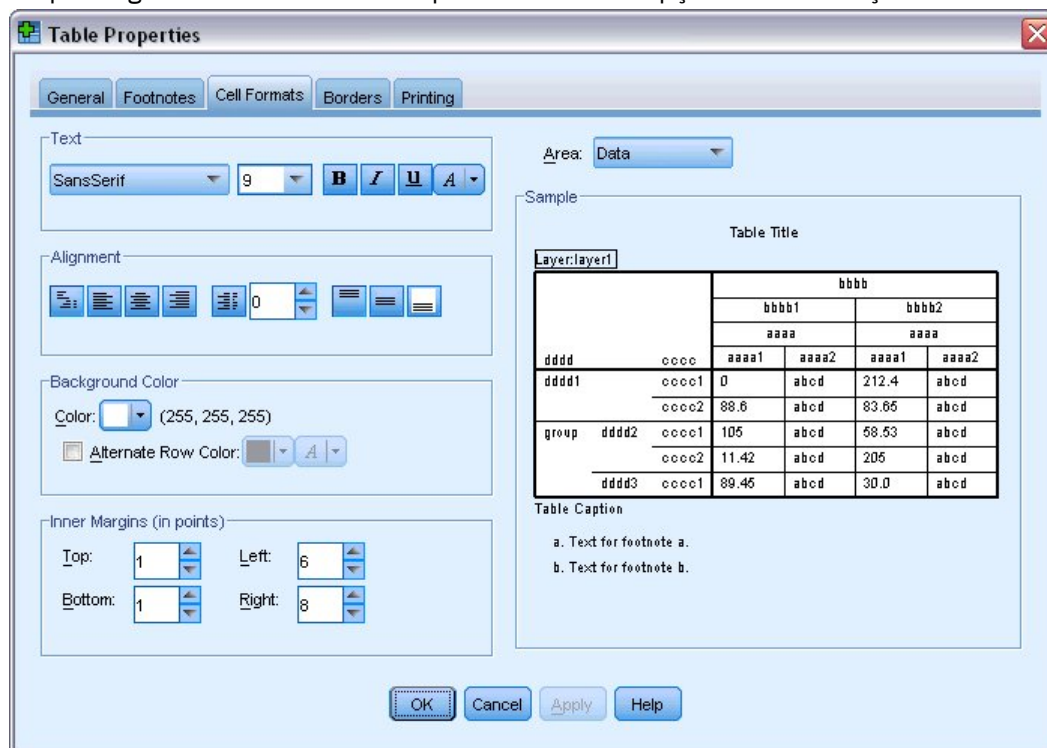


Figura 48. Caixa de diálogo Propriedades da Tabela

As opções de formatação incluem o nome da fonte, o tamanho da fonte, o estilo e a cor. Opções adicionais incluem cores de alinhamento, de texto e de plano de fundo e os tamanhos das margens.

A janela Amostra à direita fornece uma visualização de como as mudanças de formatação afetam sua tabela. Cada área da tabela pode ter estilos de formatação diferentes. Por exemplo, você provavelmente não iria querer que o título tivesse o mesmo estilo que os dados. Para selecionar uma área de tabela para editar, você pode escolher a área por nome na lista suspensa Área ou pode clicar na área que deseja mudar na janela Amostra.

5. Selecione **Dados** na lista suspensa Área.
6. Selecione uma nova cor na paleta suspensa Plano de Fundo.
7. Em seguida, selecione uma nova cor de texto.

A janela Amostra mostra o novo estilo.



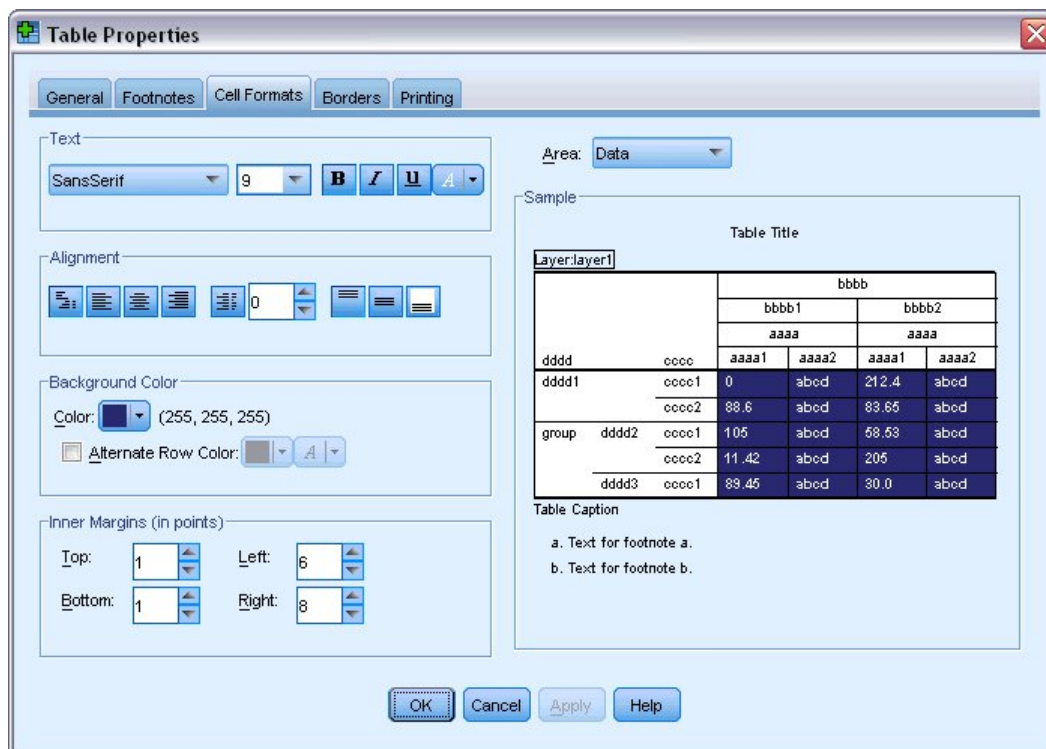


Figura 49. Mudando formatos de célula da tabela

8. Clique em **OK** para retornar à caixa de diálogo TableLooks.

É possível salvar seu novo estilo, que permite a você aplicá-lo facilmente a tabelas futuras.

9. Clique em **Salvar como**.

10. Navegue para o diretório de destino e insira um nome para seu novo estilo na caixa de texto Nome do Arquivo.

11. Clique em **Salvar**.

12. Clique em **OK** para aplicar suas mudanças e retornar ao Visualizador.

Agora, a tabela contém a formatação customizada que você especificou.

Figura 50. TableLook customizado



## 2. Clique na guia **Visualizador**.

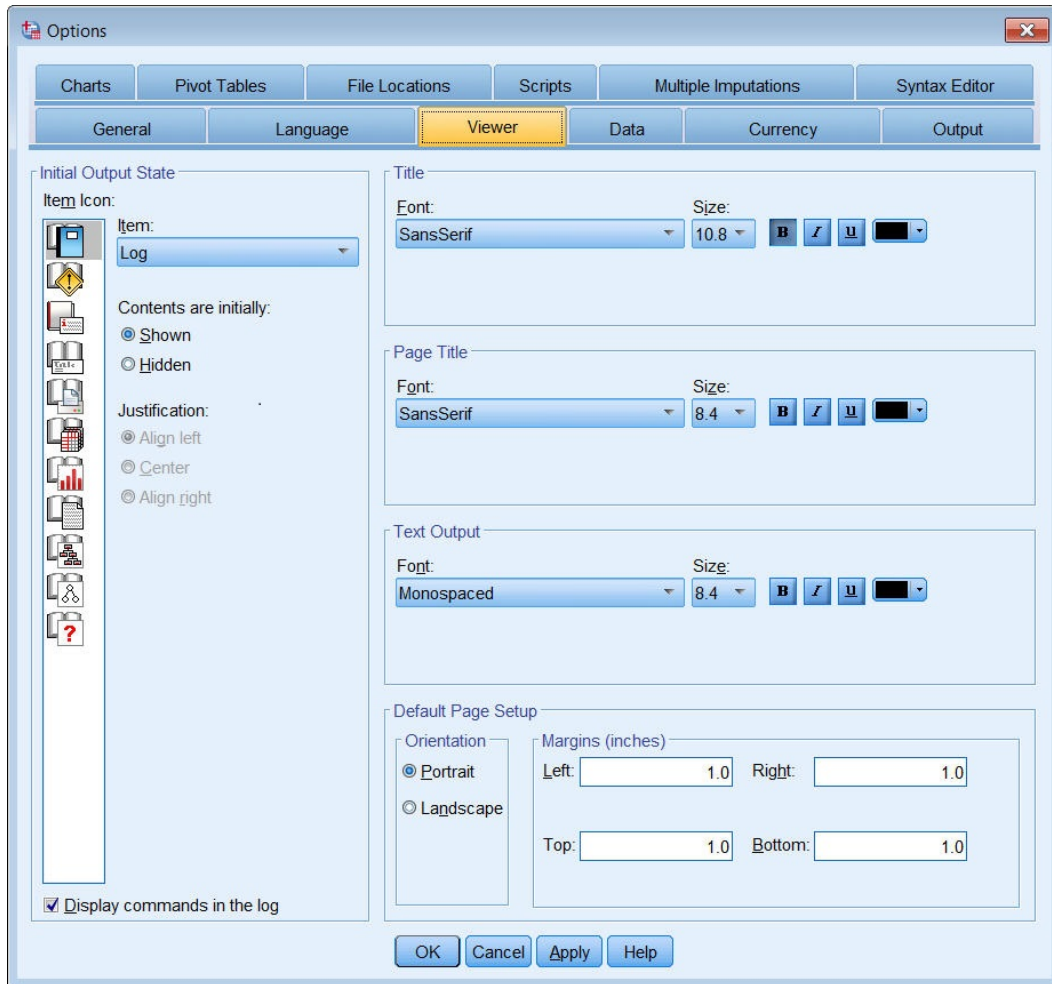


Figura 52. Opções do Visualizador

As configurações são aplicadas em uma base de objeto-por-objeto. Por exemplo, é possível customizar a maneira como os gráficos são exibidos sem fazer nenhuma mudança na maneira como as tabelas são exibidas. Simplesmente selecione o objeto que deseja customizar e faça as mudanças.

3. Clique no ícone **Título** para exibir suas configurações.
4. Clique em **Centro** para exibir todos os títulos no centro (horizontal) do Visualizador.

É possível também ocultar elementos, como as mensagens de log e de aviso, que tendem a desorganizar sua saída. Dar um clique duplo em um ícone automaticamente muda a propriedade de exibição do objeto.

5. Clique duas vezes no ícone **Avisos** para ocultar as mensagens de aviso na saída.
6. Clique em **OK** para salvar suas mudanças e fechar a caixa de diálogo.

## Exibindo rótulos de variável e de valor

Na maioria dos casos, exibir os rótulos para variáveis e valores é mais eficaz do que exibir o nome da variável ou o valor de dados reais. Pode haver casos, no entanto, em que você deseja exibir os nomes e os rótulos.

1. Nos menus, escolha:

**Editar > Opções...**

2. Clique na guia **Rótulos de Saída**.

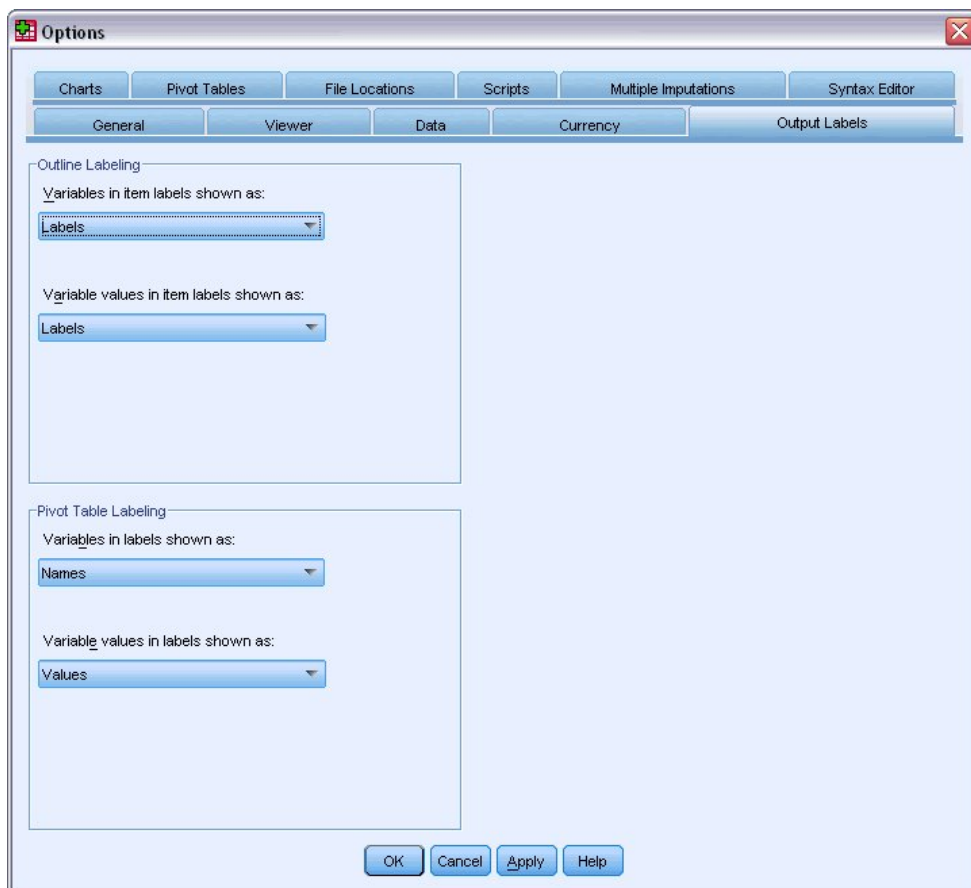


Figura 53. Configurações de rotulação de tabela dinâmica

É possível especificar configurações diferentes para as áreas de janela da estrutura de tópicos e de conteúdo. Por exemplo, para mostrar rótulos na estrutura de tópicos e nos nomes de variáveis e valores de dados no conteúdo:

3. No grupo Rotulação da Tabela Dinâmica, selecione **Nomes** nas Variáveis na lista suspensa Rótulos para mostrar nomes de variáveis em vez de rótulos.
4. Em seguida, selecione **Valores** nos Valores de Variáveis na lista suspensa Rótulos para mostra valores de dados em vez de rótulos.

Tabelas subsequentes produzidas na sessão refletirão essas mudanças.

**marital**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	3224	50.4	50.4	50.4
	1	3176	49.6	49.6	100.0
	Total	6400	100.0	100.0	

Figura 54. Nomes e valores de variáveis exibidos

## Usando resultados em outros aplicativos

Seus resultados podem ser usados em muitos aplicativos. Por exemplo, você pode querer incluir uma tabela ou gráfico em uma apresentação ou relatório.

Os exemplos a seguir são específicos para o Microsoft Word, mas eles podem funcionar de forma semelhante em outros aplicativos de processamento de palavras.

## Colando resultados como tabelas do Word

É possível colar tabelas dinâmicas no Word como tabelas nativas do Word. Todos os atributos de tabela, como tamanhos e cores das fontes, são retidos. Como a tabela é colada no formato de tabela do Word, é possível editá-la no Word como qualquer outra tabela.

1. Clique em uma tabela no Visualizador para selecioná-la.
2. Nos menus, escolha:

**Editar > Copiar**

3. Abra seu aplicativo de processamento de texto.
4. Nos menus do processador de texto, escolha:

**Editar > Colar especial...**

5. Selecione **Texto Formatado (RTF)** na caixa de diálogo Colar Especial.
6. Clique em **OK** para colar seus resultados no documento atual.

Agora, a tabela é exibida em seu documento. É possível aplicar a formatação customizada, editar os dados e redimensionar a tabela para se ajustar às suas necessidades.

## Colando resultados como texto

As tabelas dinâmicas podem ser copiadas para outros aplicativos como texto simples. Os estilos de formatação não são retidos neste método, mas você pode editar os dados da tabela depois de colá-los no aplicativo de destino.

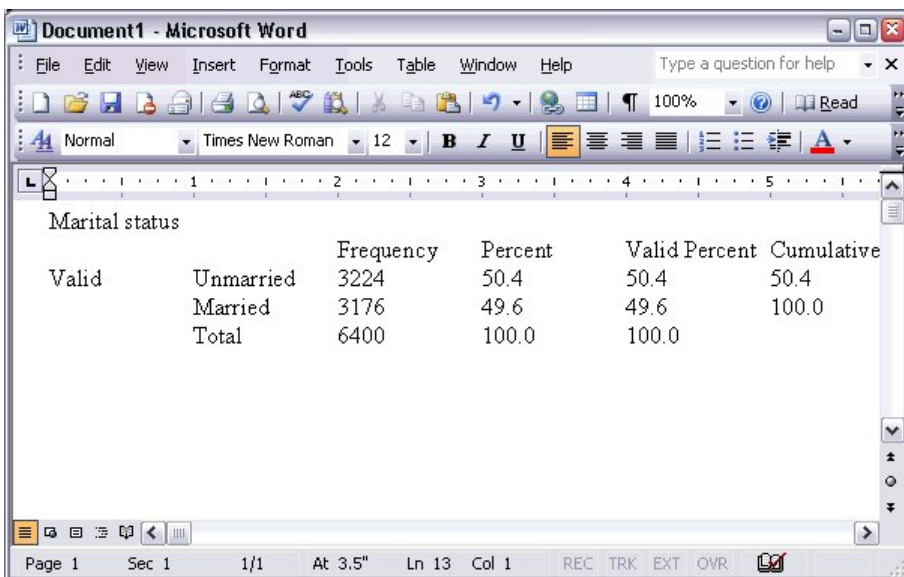
1. Clique em uma tabela no Visualizador para selecioná-la.
2. Nos menus, escolha:

**Editar > Copiar**

3. Abra seu aplicativo de processamento de texto.
4. Nos menus do processador de texto, escolha:

**Editar > Colar especial...**

5. Selecione **Texto Não Formatado** na caixa de diálogo Colar Especial.
6. Clique em **OK** para colar seus resultados no documento atual.



The screenshot shows a Microsoft Word window titled 'Document1 - Microsoft Word'. The menu bar includes File, Edit, View, Insert, Format, Tools, Table, Window, and Help. The toolbar shows various icons for document manipulation. The ribbon is set to 'Normal' with 'Times New Roman' font and size '12'. The main content area displays a table with the following data:

Marital status		Frequency	Percent	Valid Percent	Cumulative
Valid	Unmarried	3224	50.4	50.4	50.4
	Married	3176	49.6	49.6	100.0
Total		6400	100.0	100.0	

The status bar at the bottom indicates 'Page 1', 'Sec 1', '1/1', 'At 3.5"', 'Ln 13', 'Col 1', and 'REC TRK EXT OVR'.

Figura 55. Tabela dinâmica exibida no Word

Cada coluna da tabela é separada por guias. É possível mudar as larguras da coluna ajustando as paradas da tabulação no aplicativo de processamento de texto.

## Exportando resultados para arquivos do Microsoft Word, PowerPoint e Excel

É possível exportar resultados para um arquivo do Microsoft Word, PowerPoint ou Excel. É possível exportar itens selecionados ou todos os itens no Visualizador. Esta seção usa os arquivos *msouttut.spv* e *demo.sav*. Consulte o tópico [Capítulo 10, “Arquivos de Amostra”](#), na página 79 para obter mais informações

*Nota:* A exportação para PowerPoint está disponível apenas nos sistemas operacionais Windows e não está disponível com a Versão Student.

Na área de janela da estrutura de tópicos do Visualizador, você pode selecionar itens específicos que deseja exportar ou exportar todos os itens ou todos os itens visíveis.

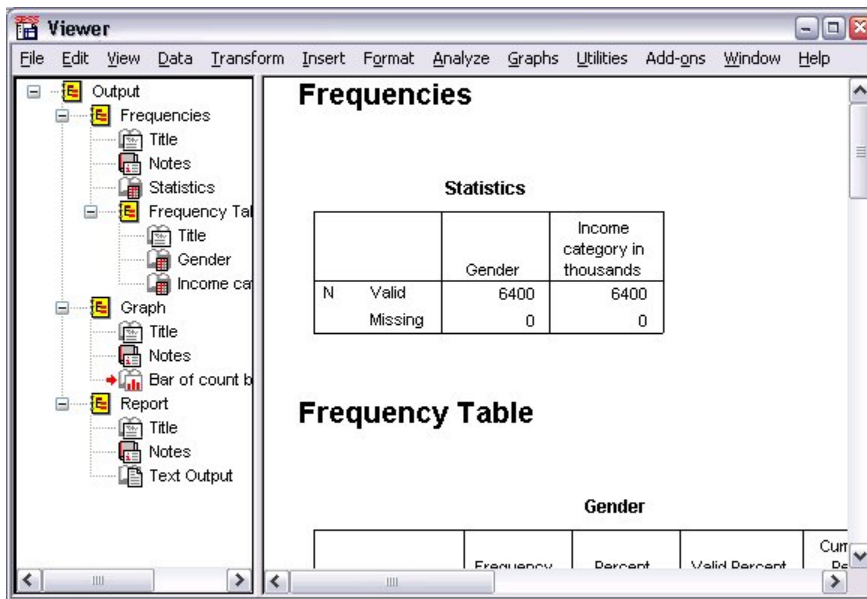


Figura 56. Visualizador

1. Nos menus do Visualizador, escolha:

### **Arquivo > Exportar...**

Em vez de exportar todos os objetos no Visualizador, você pode optar por exportar apenas os objetos visíveis (livros abertos na área de janela da estrutura de tópicos) ou aqueles que você selecionou na área de janela da estrutura de tópicos. Se não selecionou nenhum item na área de janela da estrutura de tópicos, você não tem a opção para exportar os objetos selecionados.

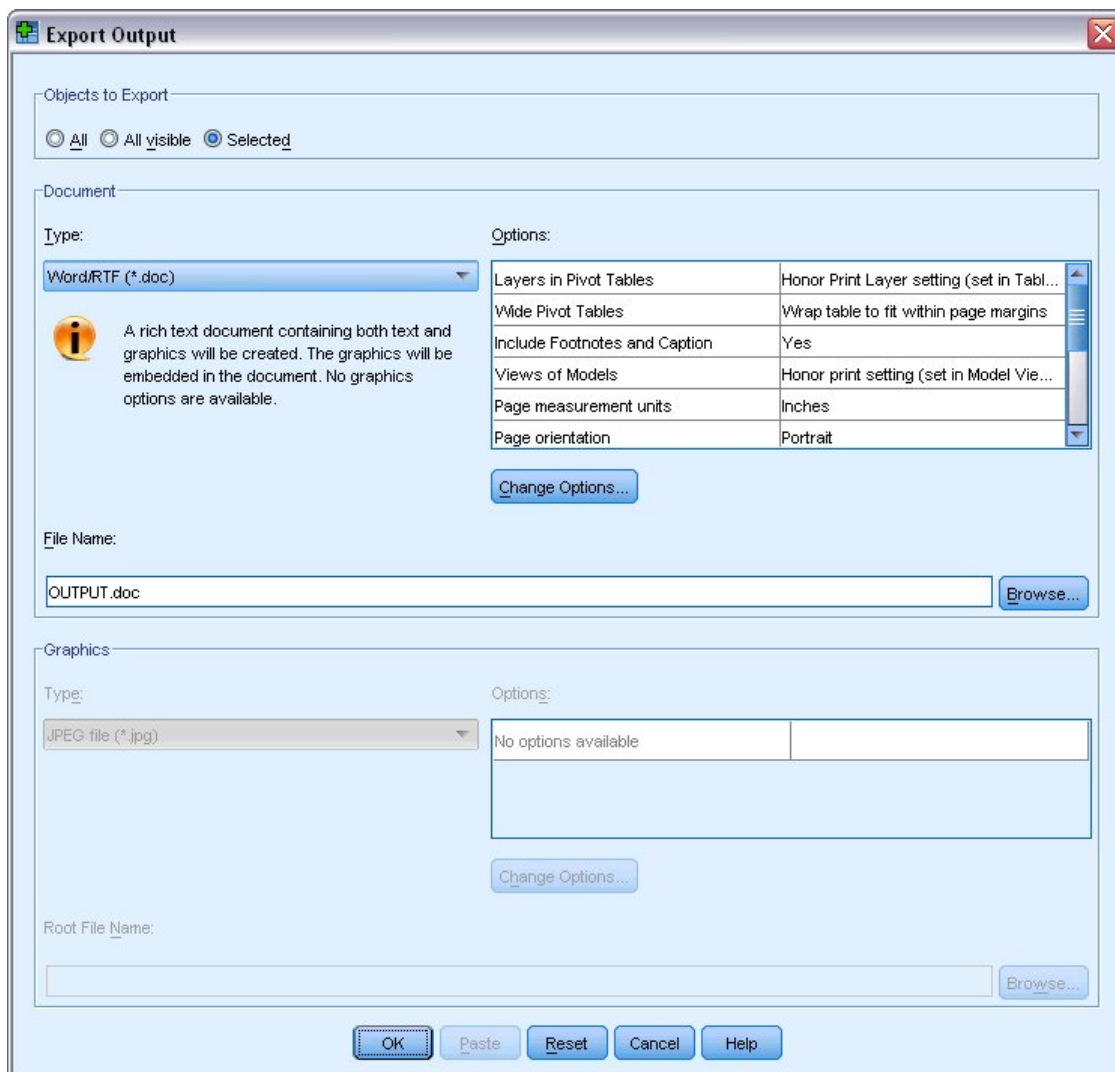


Figura 57. Caixa de diálogo Exportar Saída

2. No grupo Objetos a Exportar, selecione **Todos**.
3. Na lista suspensa Tipo, selecione **Arquivo do Word/RTF (\*.doc)**.
4. Clique em **OK** para gerar o arquivo do Word.

Ao abrir o arquivo resultante no Word, você pode ver como os resultados são exportados. As Notas, que não são objetos visíveis, aparecem no Word porque você optou por exportar todos os objetos.

As tabelas dinâmicas se tornam tabelas do Word, com toda a formatação da tabela dinâmica original retida, incluindo fontes, cores, bordas e assim por diante.

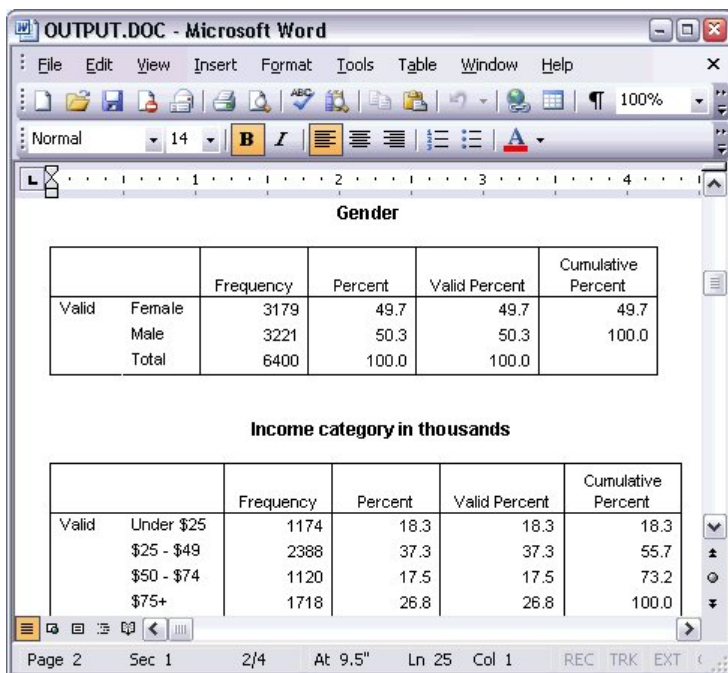


Figura 58. Tabelas dinâmicas no Word

Os gráficos são incluídos no documento do Word como imagens gráficas.

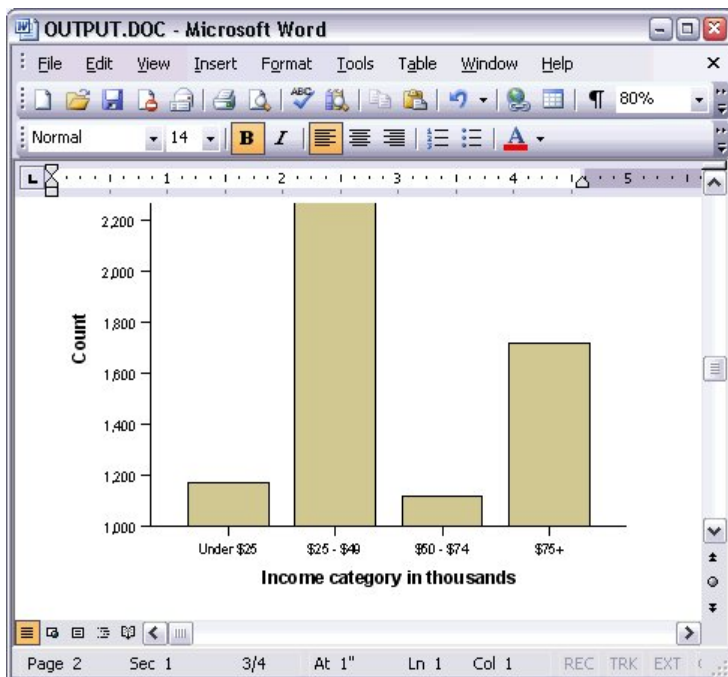


Figura 59. Gráficos no Word

A saída de texto é exibida na mesma fonte usada para o objeto de texto no Visualizador. Para um alinhamento adequado, a saída de texto deve usar uma fonte de densidade fixa (monoespaçada).



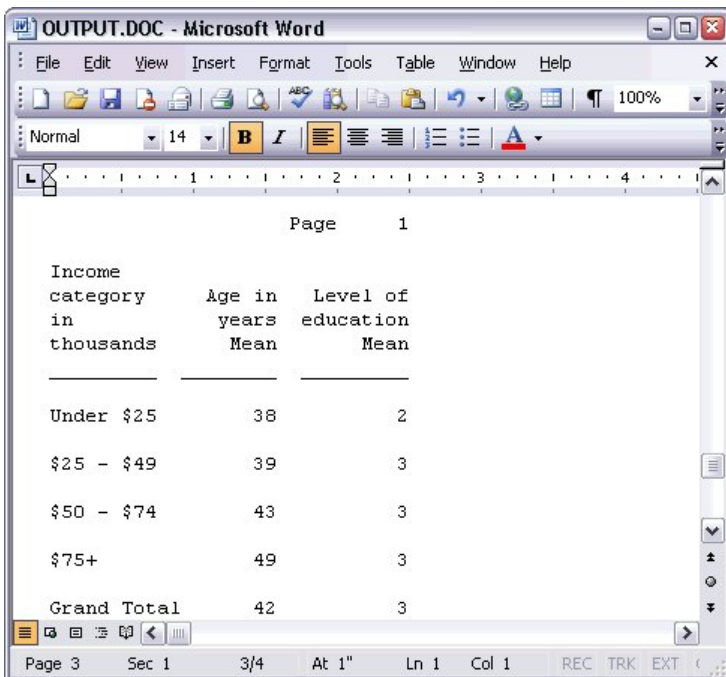


Figura 60. Saída de texto no Word

Se você exportar para um arquivo do PowerPoint, cada item exportado será colocado em um slide separado. As tabelas dinâmicas exportadas para PowerPoint se tornam tabelas do Word, com toda a formatação da tabela dinâmica original, incluindo fontes, cores, bordas e assim por diante.

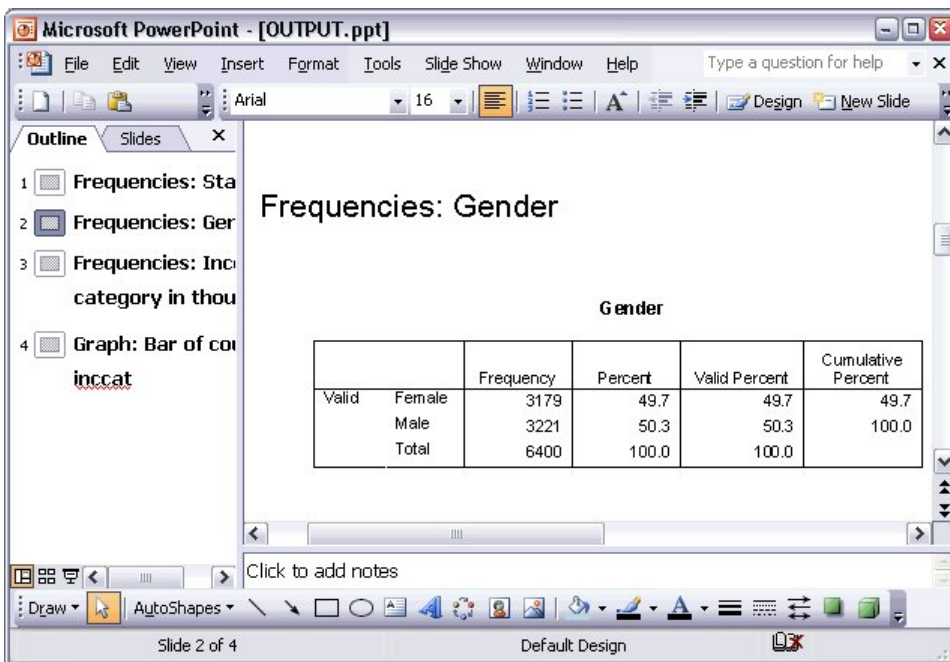


Figura 61. Tabelas dinâmicas no PowerPoint

Os gráficos selecionados para exportação para PowerPoint são integrados ao arquivo do PowerPoint.

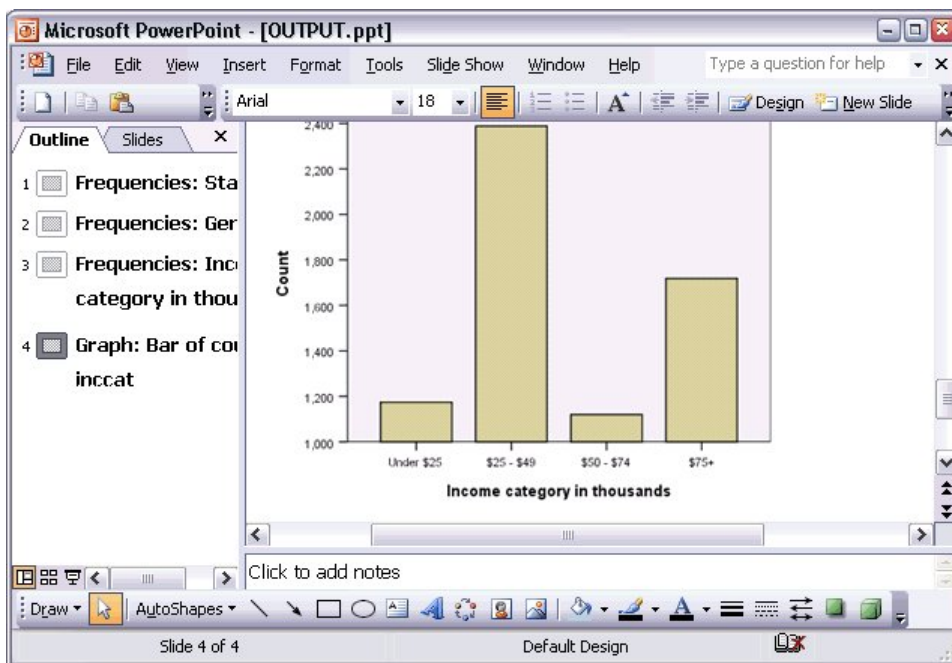


Figura 62. Gráficos no PowerPoint

Nota: A exportação para PowerPoint está disponível apenas nos sistemas operacionais Windows e não está disponível com a Versão Student.

Se você exportar para um arquivo do Excel, os resultados serão exportados de forma diferente.

The Excel output is structured as follows:

Gender					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	3,179	49.7	49.7	49.7
	Male	3,221	50.3	50.3	100.0
	Total	6,400	100.0	100.0	
Income category in thousands					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under \$25	1,174	18.3	18.3	18.3
	\$25 - \$49	2,388	37.3	37.3	55.7
	\$50 - \$74	1,120	17.5	17.5	73.2
	\$75+	1,718	26.8	26.8	100.0
	Total	6,400	100.0	100.0	

Figura 63. Output.xls no Excel

As linhas, colunas e células da tabela dinâmica se tornam linhas, colunas e células do Excel.

Cada linha na saída de texto é uma linha no arquivo do Excel, com todo o conteúdo da linha contido em uma única célula.

	A	B	C	D	E
68	Income				
69	category	Age in	Level of		
70	in	years	education		
71	thousands	Mean	Mean		
72					
73					
74	Under \$25	38	2		
75					
76	\$25 - \$49	39	3		
77					
78	\$50 - \$74	43	3		
79					
80	\$75+	49	3		
81					
82	Grand Total	42	3		

Figura 64. Saída de texto no Excel

## Exportando resultados para PDF

Você pode exportar todos os itens ou os itens selecionados no Visualizador para um arquivo PDF (Portable Document Format).

1. Nos menus na janela Visualizador que contém o resultado que você deseja exportar para PDF, escolha:

**Arquivo > Exportar...**

2. Na caixa de diálogo Exportar Saída, na lista suspensa Exportar Tipo de Arquivo de Formato, escolha **Portable Document Format**.

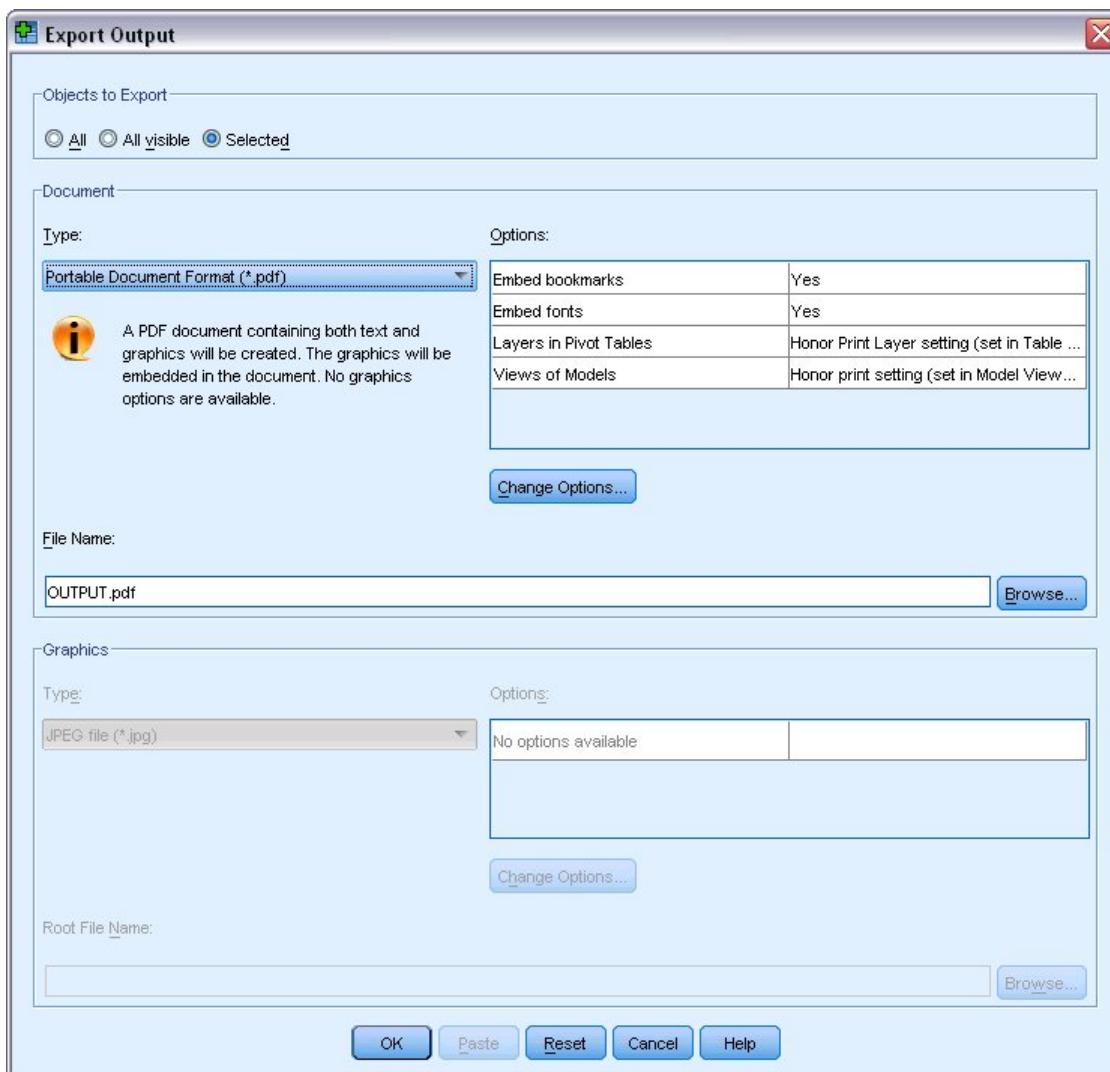


Figura 65. Caixa de diálogo Exportar Saída

- A área de janela da estrutura de tópicos do documento do Visualizador é convertida em marcadores no arquivo PDF para fácil navegação.
- O tamanho da página, a orientação, as margens, o conteúdo e a exibição dos cabeçalhos e rodapés de página e o tamanho do gráfico impresso em documentos PDF são controlados por opções de configuração de página (menu Arquivo, Configuração de Página na janela Visualizador).
- A resolução (DPI) do documento PDF é a configuração de resolução atual para a impressora padrão ou atualmente selecionada (que pode ser alterada usando Configuração de Página). A resolução máxima é 1200 DPI. Se a configuração da impressora for superior, a resolução do documento PDF será 1200 DPI. *Nota:* Documentos de alta resolução podem gerar baixos resultados quando impressos em impressoras de resolução inferior.

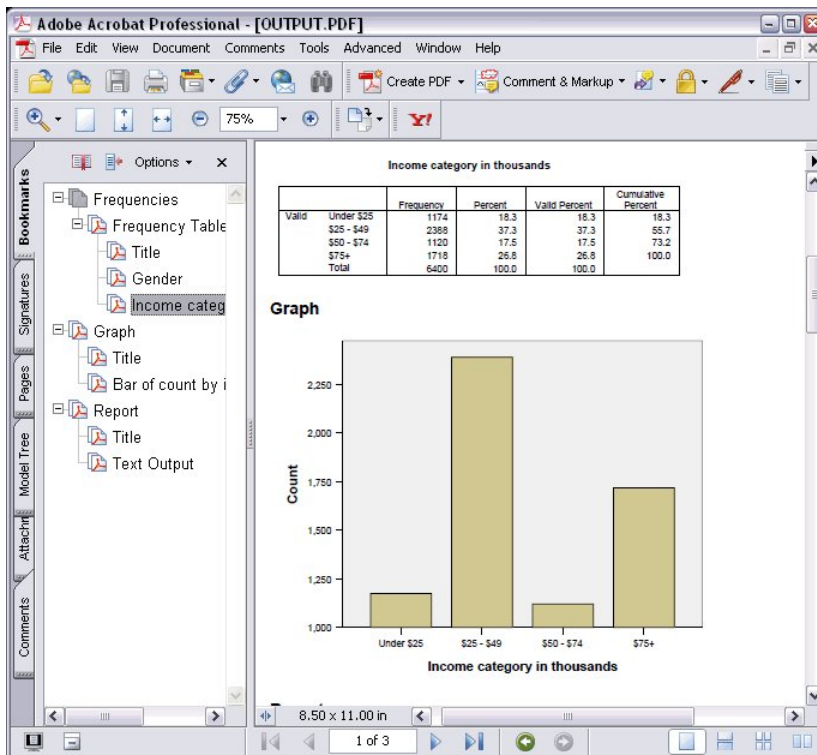


Figura 66. Arquivo PDF com marcadores

## Exportando resultados para HTML

Você também pode exportar resultados para HTML (Linguagem de Marcação de Hipertexto). Ao salvar como HTML, toda saída não gráfica é exportada em um único arquivo HTML.

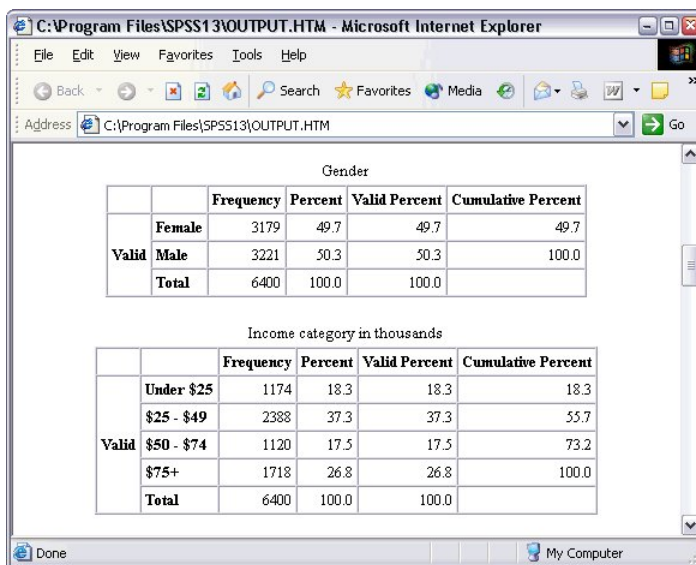


Figura 67. Output.htm no navegador da web

Quando você exporta para HTML, os gráficos podem ser exportados também, mas não para um único arquivo.

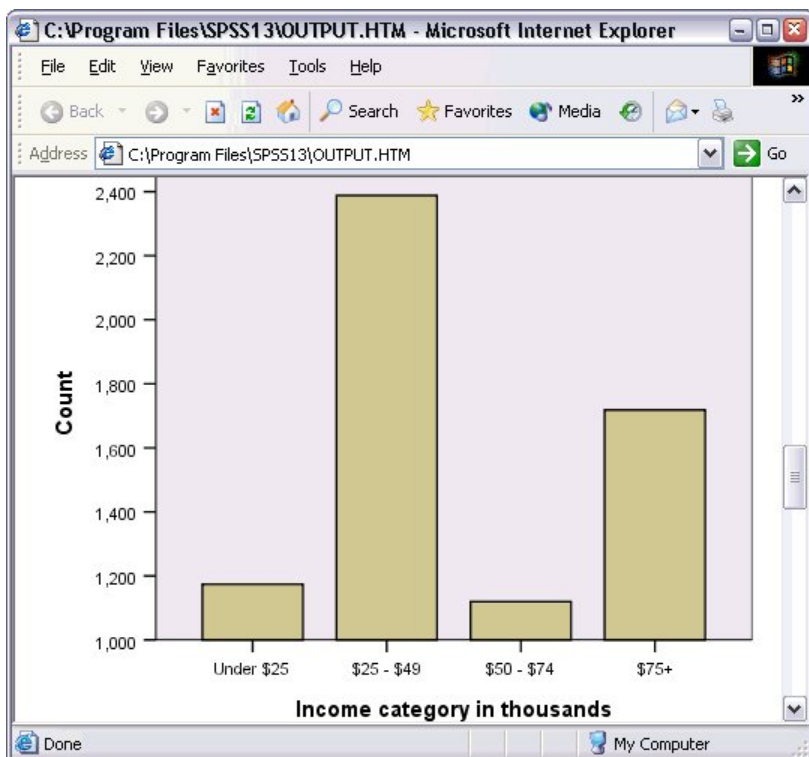


Figura 68. Gráfico em HTML

Cada gráfico será salvo como um arquivo em um formato que você especificar e as referências a esses arquivos gráficos serão colocadas no HTML. Há também uma opção para exportar todos os gráficos (ou gráficos selecionados) para arquivos gráficos separados.

## Capítulo 7. Trabalhando com sintaxe

É possível salvar e automatizar muitas tarefas comuns usando o idioma de comando poderoso. Ele também fornece algumas funcionalidades não encontradas em menus e caixas de diálogo. A maioria dos comandos é acessível a partir de menus e caixas de diálogo. No entanto, alguns comandos e opções só estão disponíveis usando o idioma de comando. O idioma de comando também permite que você salve suas tarefas em um arquivo de sintaxe para que possa repetir sua análise em uma data posterior.

Um arquivo de sintaxe de comando é simplesmente um arquivo de texto que contém comandos de sintaxe do IBM SPSS Statistics. Você pode abrir uma janela de sintaxe e digitar comandos diretamente, mas é frequentemente mais fácil deixar que as caixas de diálogo façam uma parte ou todo o trabalho para você.

Os exemplos neste capítulo usam o arquivo de dados *demo.sav*. Consulte o tópico [Capítulo 10, “Arquivos de Amostra”](#), na página 79 para obter mais informações

*Nota:* A sintaxe de comando não está disponível com a Versão Student.

### Colando sintaxe

A maneira mais fácil de criar sintaxe é usar o botão Colar localizado na maioria das caixas de diálogo.

1. Abra o arquivo de dados *demo.sav*. Consulte o tópico [Capítulo 10, “Arquivos de Amostra”](#), na página 79 para obter mais informações
2. Nos menus, escolha:  
**Analisar > Estatística Descritiva > Frequências...**
3. Selecione *Estado civil [civil]* e mova-o para a lista Variável(is).
4. Clique em **Gráficos**.
5. Na caixa de diálogo Gráficos, selecione **Gráficos de barras**.
6. No grupo Valores do Gráfico, selecione **Porcentagens**.
7. Clique em **Continuar**. Clique em **Colar** para copiar a sintaxe criada como resultado das seleções da caixa de diálogo para o Editor de Sintaxe.

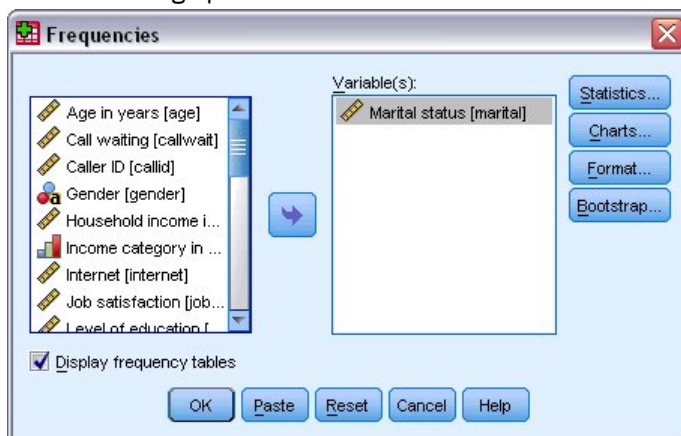


Figura 69. Caixa de diálogo Frequências

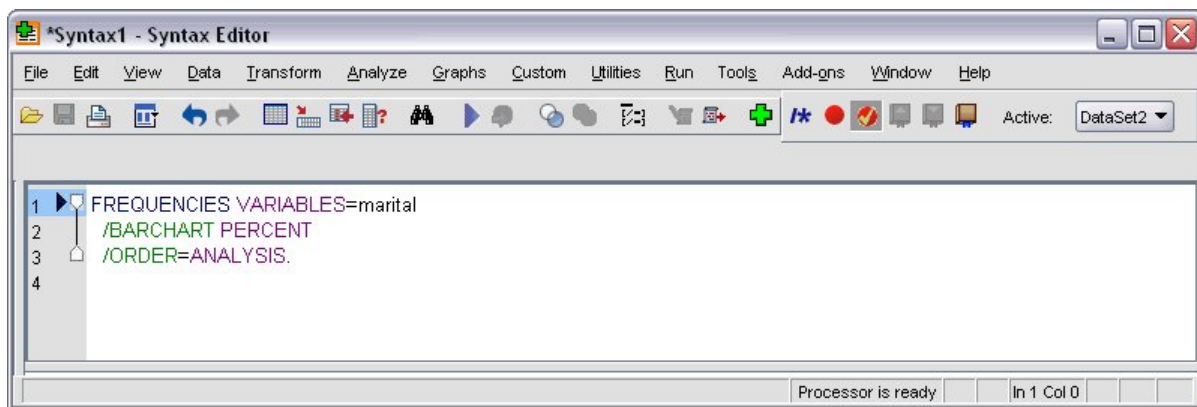


Figura 70. Sintaxe de frequências

8. Para executar a sintaxe atualmente exibida, nos menus, escolha:

**Executar > Seleção**

## Editando a sintaxe

Na janela de sintaxe, é possível editar a sintaxe. Por exemplo, você pode mudar o subcomando / BARCHART para exibir frequências em vez de porcentagens. (Um subcomando é indicado por uma barra.) Se você souber a palavra-chave para exibir frequências, pode inseri-la diretamente. Se não souber a palavra-chave, você poderá obter uma lista das palavras-chave disponíveis para o subcomando posicionando o cursor em qualquer lugar após o nome do subcomando e pressionando Ctrl+Barra de espaço. Isso exibe o controle de conclusão automática para o subcomando.

Exclua a palavra-chave PERCENTUAL do subcomando BARCHART.

Pressione Ctrl-Barra de espaço.

1. Clique no item rotulado **FREQ** para frequências. Clicar em um item no controle de conclusão automática irá inseri-lo na posição atual do cursor.

Por padrão, o controle de conclusão automática irá solicitar a você uma lista dos termos disponíveis conforme você digitar. Por exemplo, você gostaria de incluir um gráfico de pizza juntamente com o gráfico de barras. O gráfico de pizza é especificado com um subcomando separado.

2. Pressione Enter depois da palavra-chave **FREQ** e digite uma barra para indicar o início de um subcomando.

O Editor de Sintaxe solicita a você a lista de subcomandos para o comando atual.



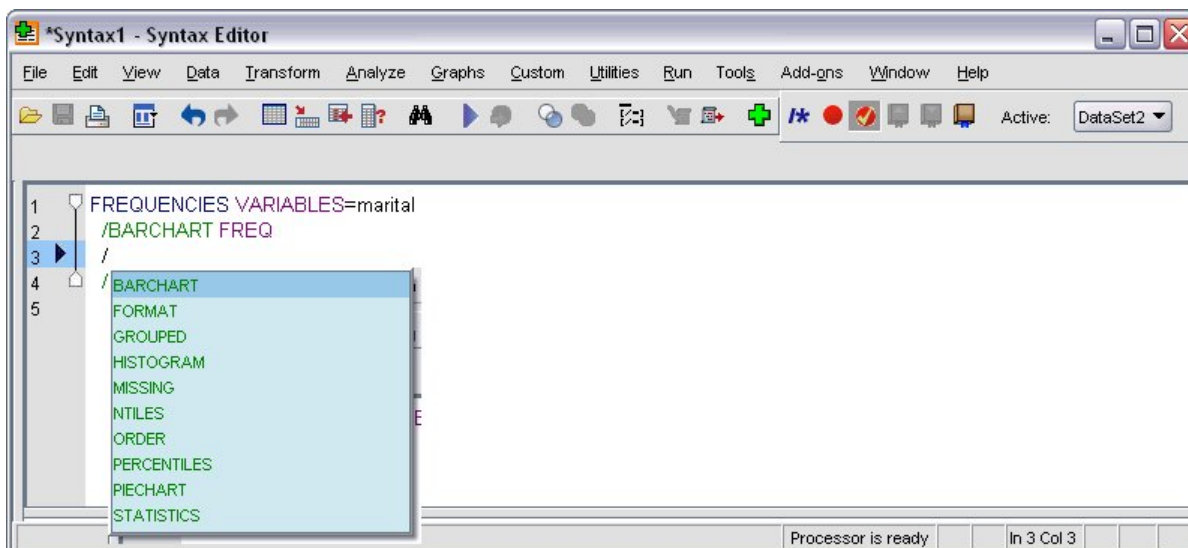


Figura 71. Controle de conclusão automática exibindo subcomandos

Para obter ajuda mais detalhada para o comando atual, pressione a tecla F1. Isso leva você diretamente às informações de referência de sintaxe de comando para o comando atual.

Você pode ter notado que o texto exibido na janela de sintaxe é colorido. A codificação de cor permite que você identifique rapidamente termos não reconhecidos, desde que apenas os termos reconhecidos sejam coloridos. Por exemplo, você digita de maneira incorreta o subcomando FORMAT como FRMAT. Os subcomandos estão na cor verde por padrão, mas o texto FRMAT aparecerá sem cor uma vez que não será reconhecido.

## Abrindo e executando um arquivo de sintaxe

1. Para abrir um arquivo de sintaxe salvo, nos menus, escolha:

**Arquivo > Abrir > Sintaxe...**

Uma caixa de diálogo padrão para abrir arquivos é exibida.

2. Selecione um arquivo de sintaxe. Se nenhum arquivo de sintaxe for exibido, certifique-se de que **Sintaxe (\*.sps)** seja selecionado como o tipo de arquivo que você deseja visualizar.

3. Clique em **Abrir**.

4. Use o menu Executar na janela de sintaxe para executar os comandos.

Se os comandos se aplicarem a um arquivo de dados específico, o arquivo de dados deverá ser aberto antes de executar os comandos ou você deverá incluir um comando que abra o arquivo de dados. Você pode colar esse tipo de comando a partir das caixas de diálogo que abrem arquivos de dados.

## Usando pontos de interrupção

Os pontos de interrupção permitem parar a execução da sintaxe de comando nos pontos especificados na sintaxe e continuar a execução quando estiver pronto. Isso permite que você visualize a saída ou os dados em um ponto intermediário em uma tarefa de sintaxe ou execute a sintaxe de comando que exibe informações sobre o estado atual dos dados, como FREQUÊNCIAS. Os pontos de interrupção podem ser configurados apenas no nível de um comando, não nas linhas específicas dentro de um comando.

Para inserir um ponto de interrupção em um comando:

1. Clique em qualquer lugar na região à esquerda do texto associado ao comando.

O ponto de interrupção é representado como um círculo vermelho na região à esquerda do texto de comando e na mesma linha que o nome do comando independentemente de onde você clicou.

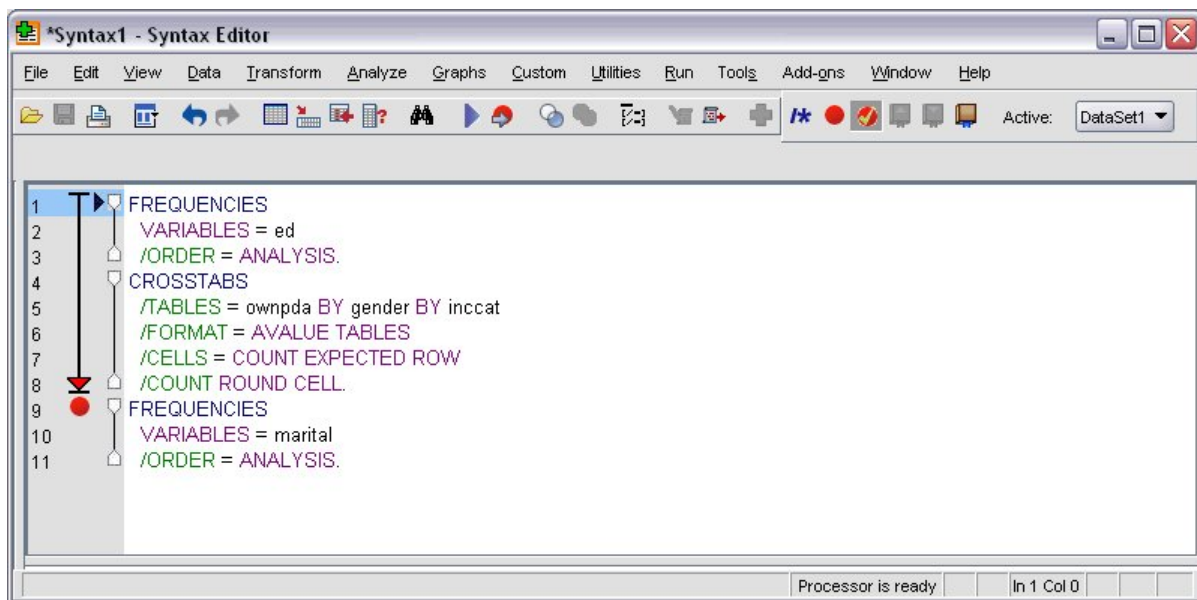


Figura 72. A execução parou em um ponto de interrupção

Quando você executa a sintaxe de comando que contém pontos de interrupção, a execução é interrompida antes de cada comando que contém um ponto de interrupção.

A seta apontando para baixo à esquerda do texto de comando mostra o progresso da execução da sintaxe. Ela abrange a região a partir do primeiro comando executado através do último comando executado e é particularmente útil ao executar a sintaxe de comando que contém pontos de interrupção.

Para retomar a execução após um ponto de interrupção:

2. Nos menus no Editor de Sintaxe, escolha:

**Executar > Continuar**

---

## Capítulo 8. Modificando valores de dados

Os dados com os quais você começa podem não ser sempre organizados da maneira mais útil para suas necessidades de análise ou de relatório. Por exemplo, você pode querer:

- Criar uma variável categórica a partir de uma variável de escala.
- Combinar várias categorias de resposta em uma única categoria.
- Criar uma nova variável que seja a diferença calculada entre duas variáveis existentes.
- Calcular o período de tempo entre duas datas.

Este capítulo usa o arquivo de dados *demo.sav*. Consulte o tópico [Capítulo 10, “Arquivos de Amostra”](#), na [página 79](#) para obter mais informações

---

### Criando uma variável categórica a partir de uma variável de escala

Várias variáveis categóricas no arquivo de dados *demo.sav* são, na verdade, derivadas de variáveis de escala nesse arquivo de dados. Por exemplo, a variável *inccat* é simplesmente *receita* agrupada em quatro categorias. Esta variável categórica usa os valores inteiros 1–4 para representar as seguintes categorias de receita (em milhares): menos de \$25, \$25–\$49, \$50–\$74 e \$75 ou mais.

Para criar a variável categórica *inccat*:

1. A partir dos menus na janela Editor de Dados, escolha:

**Transformar > Categorização Visual...**

Na caixa de diálogo Categorização Visual inicial, você seleciona a escala e/ou as variáveis ordinais para as quais deseja criar variáveis novas e categorizadas. **Categorização** significa escolher dois ou mais valores contíguos e agrupá-los na mesma categoria.

Como Categorização Visual depende de valores reais no arquivo de dados para ajudá-lo a fazer boas escolhas de categorização, é necessário ler o arquivo de dados primeiro. Como isso pode levar algum tempo se seu arquivo de dados contiver um grande número de casos, esta caixa de diálogo inicial também permite limitar o número de casos para leitura ("varredura"). Isso não é necessário para nosso arquivo de dados de amostra. Mesmo que ele contenha mais de 6.000 casos, não demora muito para varrer esse número de casos.

2. Arraste e solte *Receita doméstica em milhares [receita]* na lista Variáveis na lista Variáveis a Categorizar e, em seguida, clique em **Continuar**.

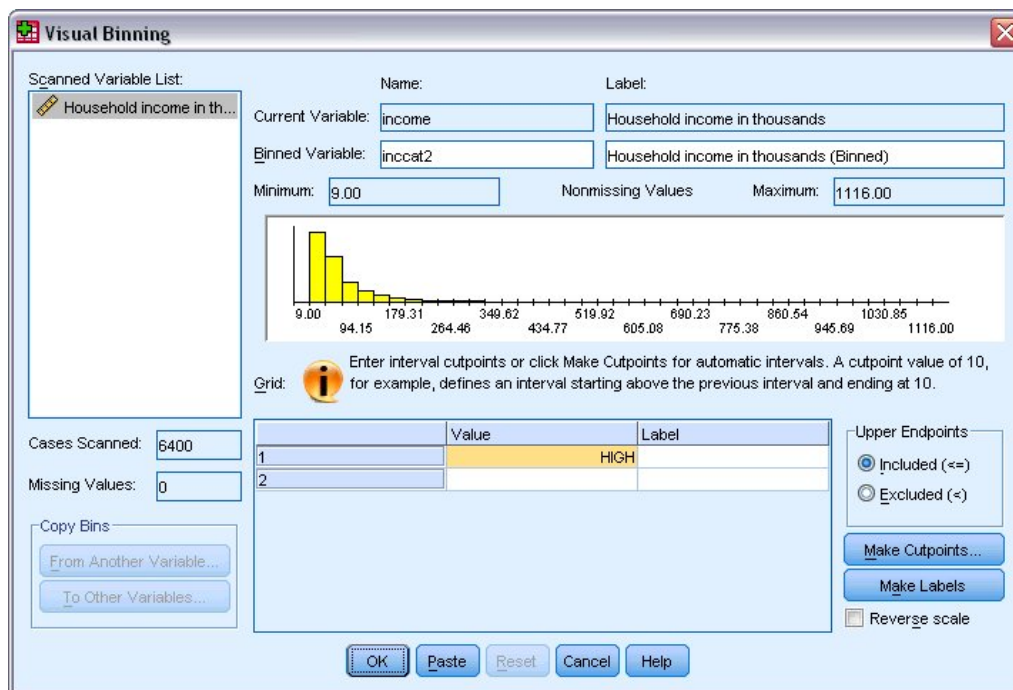


Figura 73. Caixa de diálogo Categorização Visual Principal

3. Na caixa de diálogo Categorização Visual principal, selecione *Receita doméstica em milhares [receita]* na Lista de Variáveis Varridas.

Um histograma exhibe a distribuição da variável selecionada (que, neste caso, é altamente defasada).

4. Insira *inccat2* para o novo nome de variável categorizada e *Categoria da receita [em milhares]* para o rótulo de variáveis.

5. Clique em **Fazer Pontos de Corte**.

6. Selecione **Intervalos de Largura Iguais**.

7. Insira 25 para o primeiro local do ponto de corte, 3 para o número de pontos de corte e 25 para a largura.

O número de categorias de agrupamento é um maior do que o número de pontos de corte. Portanto, neste exemplo, a nova variável categorizada terá quatro categorias, endo que cada uma das primeiras três categorias contém faixas de 25 (mil) e a última contendo todos os valores acima do valor de ponto de corte mais alto de 75 (mil).

8. Clique em **Aplicar**.

Os valores agora exibidos na grade representam os pontos de corte definidos, que são os terminais superiores de cada categoria. As linhas verticais no histograma também indicam os locais dos pontos de corte.

Por padrão, esses valores de ponto de corte são incluídos nas categorias correspondentes. Por exemplo, o primeiro valor de 25 pode incluir todos os valores menores ou iguais a 25. Mas, neste exemplo, queremos categorias que correspondam a menos de 25, 25–49, 50–74 e 75 ou mais.

9. No grupo Terminais Superiores, selecione **Excluído (<)**.

10. Em seguida, clique em **Fazer Rótulos**.

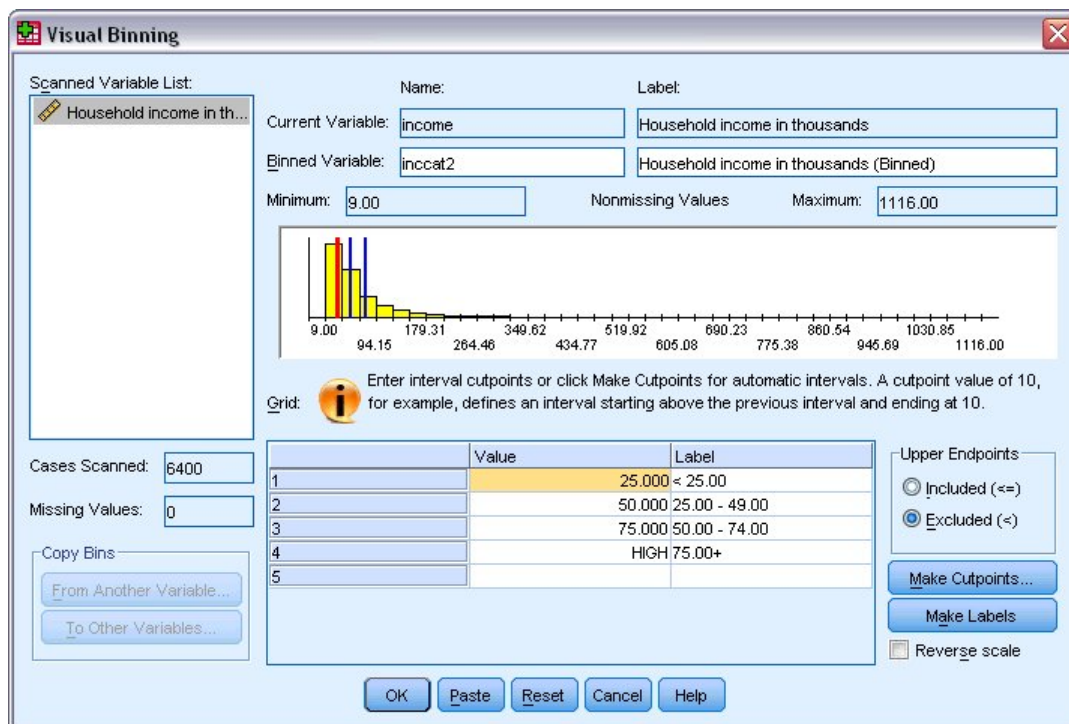


Figura 74. Rótulos de valores gerados automaticamente

Isso gera automaticamente rótulos de valores descritivos para cada categoria. Como os valores reais designados à nova variável categorizada são simplesmente números inteiros sequenciais começando com 1, os rótulos de valores podem ser muito úteis.

Você também pode inserir manualmente ou alterar pontos de corte e rótulos na grade, alterar locais de pontos de corte arrastando e soltando as linhas dos pontos de corte no histograma e excluir pontos de corte arrastando as linhas de pontos de corte para fora do histograma.

11. Clique em **OK** para criar a variável nova e categorizada.

A nova variável é exibida no Editor de Dados. Como a variável é incluída no final do arquivo, ela é exibida na coluna à extrema direita na Visualização de Dados e na última linha na Visualização de Variável.

## Calculando novas variáveis

Usando uma ampla variedade de funções matemáticas, você pode calcular novas variáveis com base em equações altamente complexas. Neste exemplo, no entanto, você irá simplesmente calcular uma nova variável que é a diferença entre os valores de duas variáveis existentes.

O arquivo de dados *demo.sav* contém uma variável para a idade atual do respondente e uma variável para o número de anos no trabalho atual. Ele não contém, no entanto, uma variável para a idade do respondente no momento em que ele ou ela começou esse trabalho. Podemos criar uma nova variável que seja a diferença calculada entre a idade atual e o número de anos no trabalho atual, que deve ser a idade aproximada em que o respondente começou esse trabalho.

1. A partir dos menus na janela Editor de Dados, escolha:

**Transformar > Calcular variável...**

2. Para Variável de Destino, insira `jobstart`.
3. Selecione *Idade em anos* [*idade*] na lista de variáveis de origem e clique no botão de seta para copiar para a caixa de texto Expressão Numérica.
4. Clique no botão de menos (–) no bloco da calculadora na caixa de diálogo (ou pressione a tecla menos no teclado).

5. Selecione *Anos com o empregador atual [empregar]* e clique no botão de seta para copiar para a expressão.

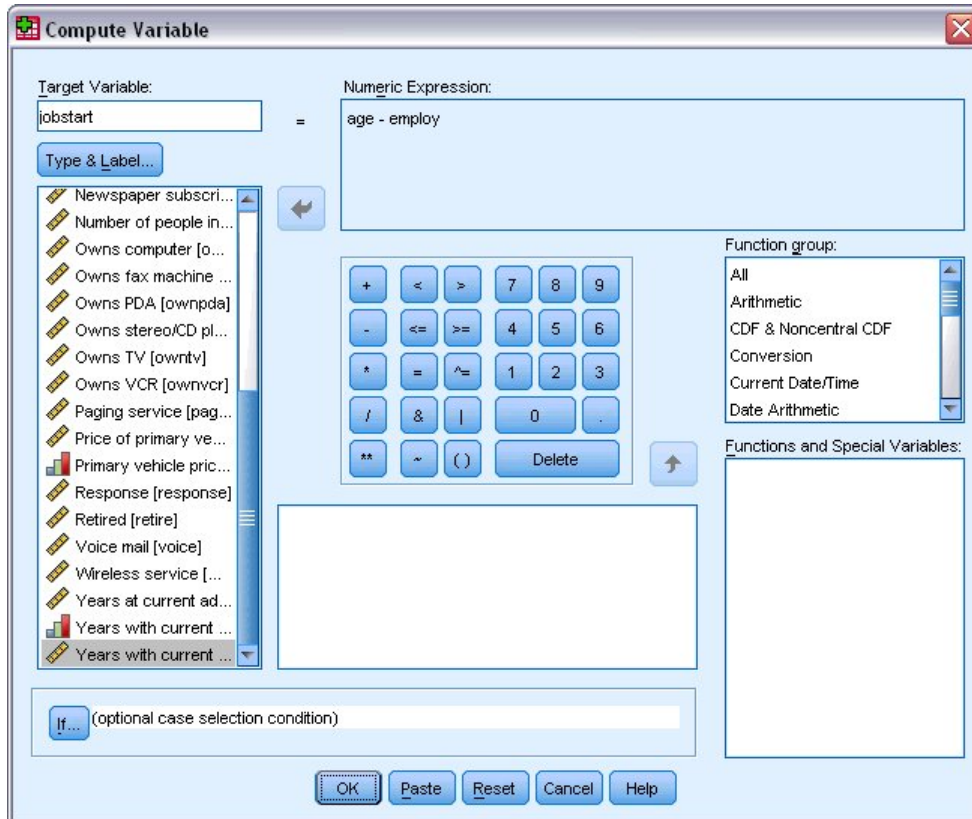


Figura 75. Caixa de diálogo Calcular Variável

*Nota:* Tenha cuidado para selecionar a variável de emprego correta. Há também uma versão categórica recodificada da variável, que *não* é o que você deseja. A expressão numérica deve ser *age-employ*, não *age-empcat*.

6. Clique em **OK** para calcular a nova variável.

A nova variável é exibida no Editor de Dados. Como a variável é incluída no final do arquivo, ela é exibida na coluna à extrema direita na Visualização de Dados e na última linha na Visualização de Variável.

## Usando funções em expressões

Você também pode usar funções predefinidas em expressões. Mais de 70 funções integradas estão disponíveis, incluindo:

- Funções aritméticas
- Funções estatísticas
- Funções de distribuição
- Funções lógicas
- Funções de agregação e extração de data e hora
- Funções de valor omissos
- Funções de caso cruzado
- Funções de sequência de caracteres

As funções são organizadas em grupos logicamente distintos, como um grupo para operações aritméticas e outro para calcular métricas estatísticas. Por conveniência, várias variáveis de sistema comumente usadas, como *\$TIME* (data e hora atuais), também são incluídas nos grupos de função apropriados.

Colando uma função em uma expressão

Para colar uma função em uma expressão:

1. Posicione o cursor na expressão no ponto em que deseja que a função apareça.
2. Selecione o grupo apropriado na lista de grupos Função. O grupo rotulado **Tudo** fornece uma lista de todas as funções e variáveis de sistema disponíveis.
3. Clique duas vezes na função na lista Funções e Variáveis Especiais (ou selecione a função e clique na seta adjacente à lista de grupos Função).

A função é inserida na expressão. Se você destacar parte da expressão e, em seguida, inserir a função, a parte destacada da expressão será usada como o primeiro argumento na função.

Editando uma Função em uma Expressão

A função não é concluída até você inserir os argumentos, representados por pontos de interrogação na função colada. O número de pontos de interrogação indica o número mínimo de argumentos requeridos para concluir a função.

1. Destaque o(s) ponto(s) de interrogação na função colada.
2. Insira os argumentos. Se os argumentos forem nomes de variáveis, você poderá colá-los a partir da lista de variáveis.

## Usando expressões condicionais

Você pode usar expressões condicionais (também chamadas de expressões lógicas) para aplicar transformações aos subconjuntos selecionados de casos. Uma expressão condicional retorna um valor de verdadeiro, falso ou omissivo para cada caso. Se o resultado de uma expressão condicional for verdadeiro, a transformação será aplicada a esse caso. Se o resultado for falso ou omissivo, a transformação não será aplicada ao caso.

Para especificar uma expressão condicional:

1. Clique em **Se** na caixa de diálogo Calcular Variável. Isso abre a caixa de diálogo Casos Se.
2. Selecione **Incluir o caso se satisfaz a condição**.
3. Insira a expressão condicional.

A maioria das expressões condicionais contém pelo menos um operador relacional, como em:

```
age>=21
```

ou

```
income*3<100
```

No primeiro exemplo, apenas os casos com um valor de 21 ou maior para *Idade [idade]* são selecionados. No segundo exemplo, *Receita doméstica em milhares [receita]* multiplicado por 3 deve ser menor que 100 para que um caso seja selecionado.

Você também pode vincular duas ou mais expressões condicionais usando operadores lógicos, como em:

```
age>=21 | ed>=4
```

ou

```
income*3<100 & ed=5
```

No primeiro exemplo, os casos que atendem à condição *Idade [idade]* ou à condição *Nível de educação [ed]* é selecionada. No segundo exemplo, ambas as condições, *Receita doméstica em milhares [receita]* e *Nível de educação [ed]*, devem ser atendidas para que um caso seja selecionado.

## Trabalhando com datas e horas

Várias tarefas comumente executadas com datas e horas podem ser facilmente realizadas usando-se o Assistente de data e hora. Usando este assistente, você pode:

- Criar uma variável de data/hora a partir de uma variável de sequência que contém uma data ou hora.

- Construir uma variável de data/hora mesclando variáveis que contêm partes diferentes da data ou hora.
- Incluir ou subtrair valores de variáveis de data/hora, incluindo adicionar ou subtrair duas variáveis de data/hora.
- Extrair uma parte de uma variável de data ou hora; por exemplo, o dia do mês a partir de uma variável de data/hora que tem o formato mm/dd/yyyy.

Os exemplos nesta seção usam o arquivo de dados *upgrade.sav*. Consulte o tópico [Capítulo 10, “Arquivos de Amostra”](#), na página 79 para obter mais informações

Para usar o Assistente de data e hora:

1. Nos menus, escolha:

### Transformação > Assistente de data e hora...



Figura 76. Tela de introdução do Assistente de Data e Hora

A tela de introdução do Assistente de Data e Hora apresenta a você um conjunto de tarefas gerais. As tarefas que não se aplicam aos dados atuais são desativadas. Por exemplo, o arquivo de dados *upgrade.sav* não contém nenhuma variável de sequência, portanto, a tarefa para criar uma variável de data a partir de uma sequência é desativada.

Se você não estiver familiarizado com datas e horas no IBM SPSS Statistics, poderá selecionar **Aprender como datas e horas são representadas** e clique em **Avançar**. Isso leva a uma tela que fornece uma breve visão geral das variáveis de data/hora e um link, através do botão Ajuda, para informações mais detalhadas.

## Calculando o período de tempo entre duas datas

Uma das tarefas mais comuns envolvendo datas é calcular o período de tempo entre duas datas. Como exemplo, considere uma empresa de software interessada em analisar as compras de licenças de atualização determinando o número de anos desde que cada cliente comprou uma atualização pela última vez. O arquivo de dados *upgrade.sav* contém uma variável para a data em que cada cliente comprou uma atualização pela última vez, mas não o número de anos desde essa compra. Uma nova variável que é o período de tempo em anos entre a data da última atualização e a data da próxima liberação do produto fornecerá uma medida dessa quantidade.

Para calcular o período de tempo entre duas datas:



1. Selecione **Calcular com datas e horas** na tela de introdução do Assistente de Data e Hora e clique em **Avançar**.
2. Selecione **Calcular o número de unidades de tempo entre duas datas** e clique em **Avançar**.

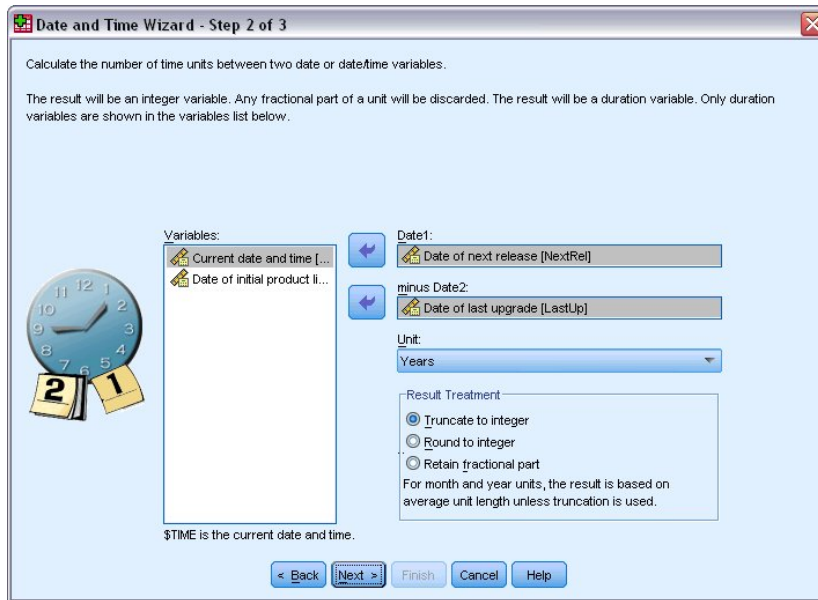


Figura 77. Calculando o período de tempo entre duas datas: Etapa 2

3. Na etapa 2, selecione *Data da próxima liberação* para Date1.
4. Selecione *Data da última atualização* para Date2.
5. Selecione **Anos** para a Unidade e **Truncar para Número Inteiro** para o Tratamento de Resultado. (Estas são as seleções padrão.)
6. Clique em **Avançar**.
7. Na etapa 3, insira *YearsLastUp* para o nome da variável de resultado. As variáveis de resultado não podem ter o mesmo nome que uma variável existente.
8. Insira *Anos desde a última atualização* como o rótulo para a variável de resultado. Os rótulos de variáveis para variáveis de resultado são opcionais.
9. Deixe a seleção padrão de **Criar a variável agora** e clique em **Concluir** para criar a nova variável.

A nova variável, *YearsLastUp*, exibida no Editor de Dados é o número inteiro de anos entre as duas datas. Partes fracionárias de um ano foram truncadas.

## Incluindo uma duração para uma data

Você pode incluir ou subtrair durações, como 10 dias ou 12 meses, para uma data. Continuando com o exemplo da empresa de software da seção anterior, considere determinar a data em que o contrato de suporte técnico inicial de cada cliente termina. O arquivo de dados *upgrade.sav* contém uma variável para o número de anos de suporte contratado e uma variável para a data de compra inicial. É possível, então, determinar a data de encerramento do suporte inicial incluindo anos de suporte na data de compra.

Para incluir uma duração em uma data:

1. Selecione **Calcular com datas e horas** na tela de introdução do Assistente de Data e Hora e clique em **Avançar**.
2. Na etapa 1, selecione **Incluir ou subtrair uma duração de uma data** e clique em **Avançar**.

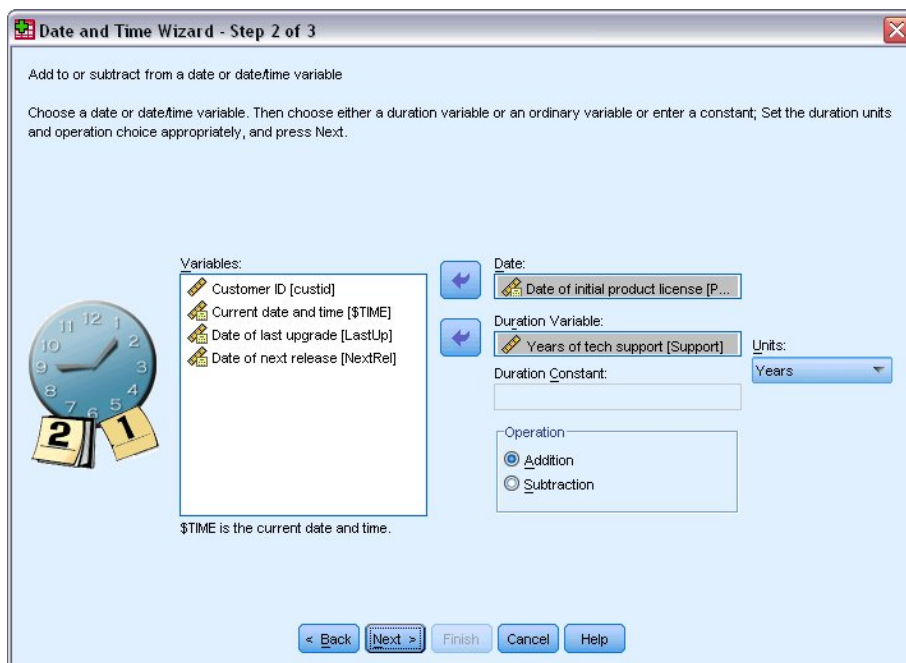


Figura 78. Incluindo uma duração em uma data: Etapa 2

3. Selecione *Data da licença de produto inicial* para Data.
  4. Na etapa 2, selecione *Anos de suporte técnico* para a Variável de Duração.  
Como *Anos de suporte técnico* é simplesmente uma variável numérica, você precisa indicar as unidades a serem usadas ao incluir esta variável como uma duração.
  5. Selecione **Anos** na lista suspensa Unidades.
  6. Clique em **Avançar**.
  7. Na etapa 3, insira *SupEndDate* para o nome da variável de resultado. As variáveis de resultado não podem ter o mesmo nome que uma variável existente.
  8. Insira *Data de encerramento para suporte* como o rótulo para a variável de resultado. Os rótulos de variáveis para variáveis de resultado são opcionais.
  9. Clique em **Concluir** para criar a nova variável.
- A nova variável é exibida no Editor de Dados.

---

## Capítulo 9. Classificando e selecionando dados

Os arquivos de dados não são sempre organizados no formato ideal para suas necessidades específicas. Para preparar dados para análise, você pode selecionar a partir de uma ampla faixa de transformações de arquivo, incluindo a capacidade para:

- **Classificar dados.** É possível classificar casos com base no valor de uma ou mais variáveis.
- **Selecionar subconjuntos de casos.** É possível restringir sua análise a um subconjunto de casos ou executar análises simultâneas em subconjuntos diferentes.

Os exemplos neste capítulo usam o arquivo de dados *demo.sav*. Consulte o tópico [Capítulo 10, “Arquivos de Amostra”](#), na página 79 para obter mais informações

---

### Classificando Dados

A classificação de casos (classificação de linhas do arquivo de dados) é frequentemente útil e às vezes necessária para determinados tipos de análise.

Para reordenar a sequência de casos no arquivo de dados com base no valor de uma ou mais variáveis de classificação:

1. Nos menus, escolha:

**Dados > Classificar casos...**

A caixa de diálogo Classificar Casos é exibida.

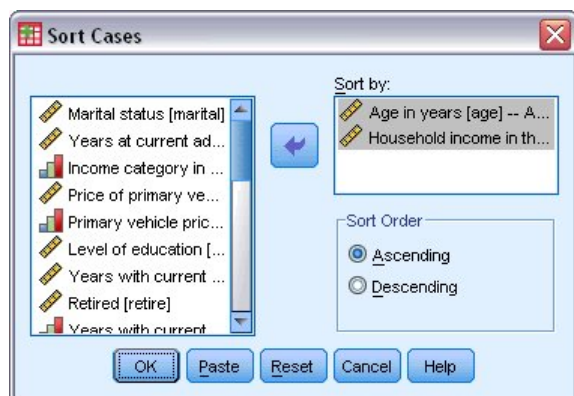


Figura 79. Caixa de diálogo Classificar Casos

2. Inclua as variáveis *Idade em anos [idade]* e *Receita familiar em milhares [receita]* na lista Classificar por.

Se você selecionar várias variáveis de classificação, a ordem em que elas aparecem na lista Classificar por determina a ordem na qual os casos são classificados. Neste exemplo, com base nas entradas na lista Classificar por, os casos serão classificados pelo valor de *Receita familiar em milhares [receita]* dentro das categorias de *Idade em anos [idade]*. Para variáveis de sequência, letras maiúsculas precedem suas contrapartes minúsculas na ordem de classificação (por exemplo, o valor de sequência *Sim* vem antes de *sim* na ordem de classificação).

---

### Processamento de divisão de arquivo

Para dividir seu arquivo de dados em grupos separados para análise:

1. Nos menus, escolha:

**Dados > Dividir arquivo...**

A caixa de diálogo Dividir Arquivo é exibida.

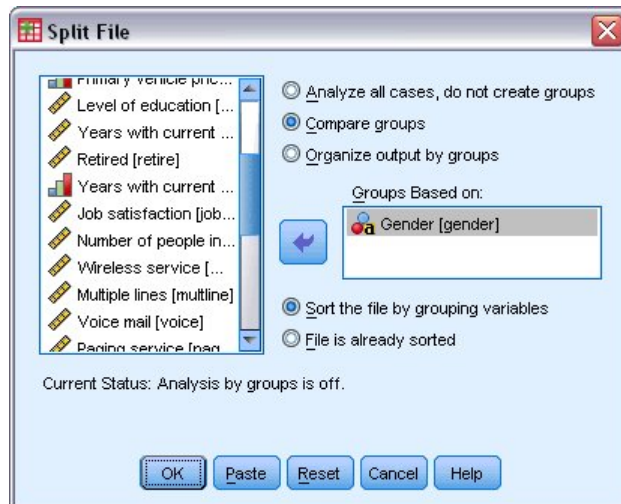


Figura 80. Caixa de diálogo Dividir Arquivo

2. Selecione **Comparar grupos** ou **Organizar saída por grupos**. (Os exemplos que vêm depois dessas etapas mostram as diferenças entre essas duas opções.)
3. Selecione *Sexo [sexo]* para dividir o arquivo em grupos separados para essas variáveis.

É possível usar variáveis numéricas, de sequência curta e de sequência longa como variáveis de agrupamento. Uma análise separada é executada por cada subgrupo que é definido pelas variáveis de agrupamento. Se você selecionar várias variáveis de agrupamento, a ordem na qual elas aparecem na lista Grupos Baseados em determina a maneira em que os casos são agrupados.

Se você selecionar **Comparar grupos**, os resultados de todos os grupos de divisão de arquivo serão incluídos na(s) mesma(s) tabela(s), conforme mostrado na tabela de estatísticas de resumo a seguir que é gerada pelo procedimento Frequências.

**Statistics**

Household income in thousands

Female	N	Valid	3179
		Missing	0
	Mean		68.7798
	Median		44.0000
	Std. Deviation		75.73510
Male	N	Valid	3221
		Missing	0
	Mean		70.1608
	Median		45.0000
	Std. Deviation		81.56216

Figura 81. Saída de divisão de arquivo com tabela dinâmica única

Se você selecionar **Organizar saída por grupos** e executar o procedimento Frequências, duas tabelas dinâmicas serão criadas: uma tabela para mulheres e uma tabela para homens.

**Statistics<sup>a</sup>**

Household income in thousands

N	Valid	3179
	Missing	0
Mean		68.7798
Median		44.0000
Std. Deviation		75.73510

a. Gender = Female

Figura 82. Saída de divisão de arquivo com tabela dinâmica para mulheres

**Statistics<sup>a</sup>**

Household income in thousands

N	Valid	3221
	Missing	0
Mean		70.1608
Median		45.0000
Std. Deviation		81.56216

a. Gender = Male

Figura 83. Saída de divisão de arquivo com tabela dinâmica para homens

## Classificando casos para o processamento de divisão de arquivo

O procedimento Dividir Arquivo cria um novo subgrupo toda vez que encontra um valor diferente para uma das variáveis de agrupamento. Portanto, é importante classificar casos com base nos valores das variáveis de agrupamento antes de chamar o processamento de divisão de arquivo.

Por padrão, Dividir Arquivo automaticamente classifica o arquivo de dados com base nos valores das variáveis de agrupamento. Se o arquivo já estiver classificado na ordem adequada, você poderá economizar o tempo de processamento se selecionar **O arquivo já está classificado**.

## Ligando e desligando o processamento de divisão de arquivo

Depois de você chamar o processamento de divisão de arquivo, ele permanece em vigor para o restante da sessão, a menos que você o desligue.

- **Analisar todos os casos.** Esta opção desliga o processamento de divisão de arquivo.
- **Comparar grupos e Organizar saída por grupos.** Esta opção liga o processamento de divisão de arquivo.

Se o processamento de divisão de arquivo estiver em vigor, a mensagem **Dividir arquivo em** aparecerá na barra de status na parte inferior da janela de aplicativo.

## Selecionando subconjuntos de casos

É possível restringir sua análise a um subgrupo específico com base nos critérios que incluem variáveis e expressões complexas. É possível também selecionar uma amostra aleatória de casos. Os critérios usados para definir um subgrupo podem incluir:

- Valores de variáveis e intervalos
- Intervalos de data e hora
- Números de casos (linhas)
- Expressões aritméticas
- Expressões Lógicas
- Funções

Para selecionar um subconjunto de casos para análise:

1. Nos menus, escolha:

**Dados > Selecionar Casos...**

Isso abre a caixa de diálogo Selecionar Casos.

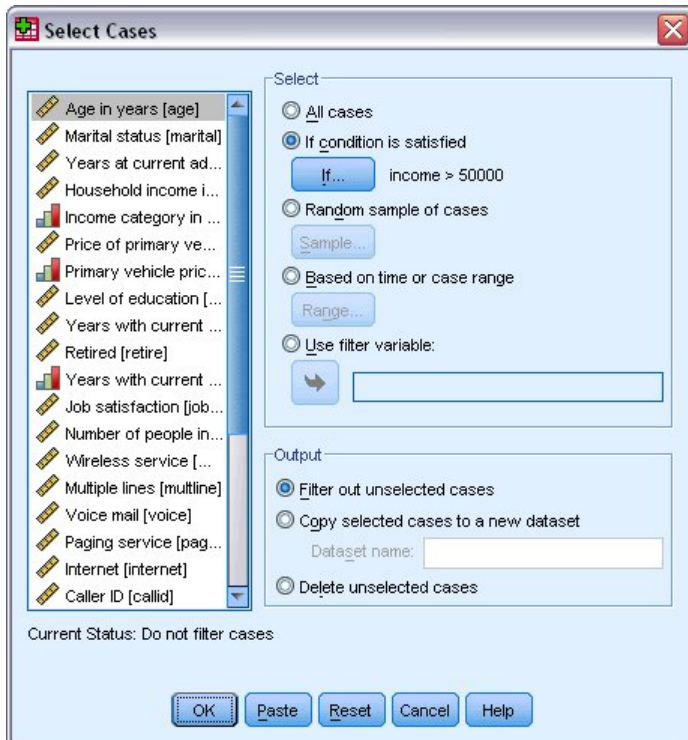


Figura 84. Caixa de diálogo Selecionar Casos

## Selecionando casos com base em expressões condicionais

Para selecionar casos com base em uma expressão condicional:

1. Selecione **Se a condição for satisfeita** e clique em **Se** na caixa de diálogo Selecionar Casos.

Isso abre a caixa de diálogo Selecionar Casos Se.

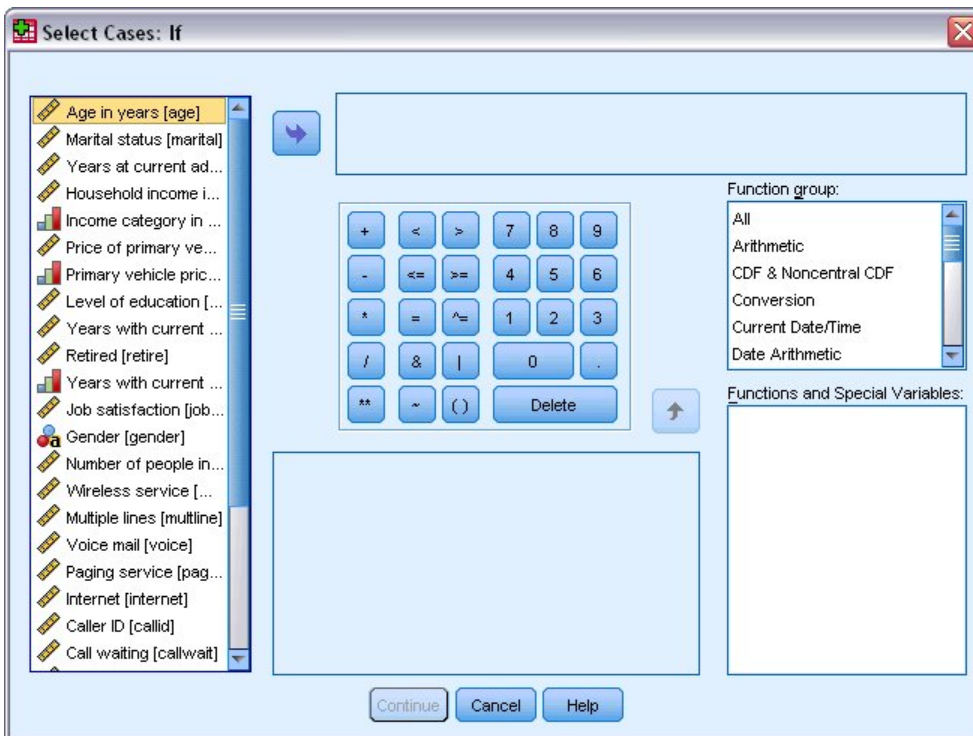


Figura 85. Caixa de diálogo Selecionar Casos Se

A expressão condicional pode usar nomes de variáveis existentes, constantes, operadores aritméticos, operadores lógicos, operadores relacionais e funções. É possível digitar e editar a expressão na caixa de texto apenas como texto em uma janela de saída. É possível também usar o teclado da calculadora, a lista de variáveis e a lista de funções para colar elementos na expressão. Consulte o tópico “[Usando expressões condicionais](#)” na página 67 para obter mais informações

## Selecionando uma amostra aleatória

Para obter uma amostra aleatória:

1. Selecione **Amostra aleatória de casos** na caixa de diálogo Selecionar Casos.
2. Clique em **Amostra**.

Isso abre a caixa de diálogo Selecionar Amostra Aleatória de Casos.

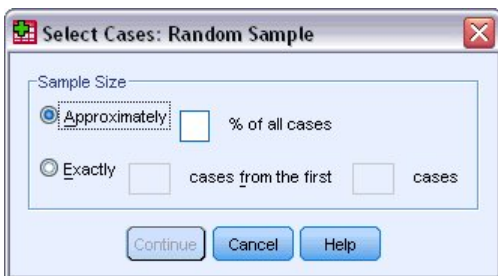


Figura 86. Caixa de diálogo Selecionar Amostra Aleatória de Casos

É possível selecionar uma das alternativas a seguir para o tamanho da amostra:

- **Aproximadamente.** Uma porcentagem especificada pelo usuário. Esta opção gera uma amostra aleatória de aproximadamente a porcentagem especificada de casos.
- **Exatamente.** Um número de casos especificado pelo usuário. Você deve também especificar o número de casos a partir dos quais gerar a amostra. Este segundo número deve ser menor ou igual ao número

total de casos no arquivo de dados. Se o número exceder o número total de casos no arquivo de dados, a amostra conterá proporcionalmente menos casos do que o número solicitado.

## Selecionando um intervalo de tempo ou um intervalo de caso

Para selecionar um intervalo de casos com base em datas, horas ou números de observações (linhas):

1. Selecione **Baseado no intervalo de hora ou de caso** e clique em **Intervalo** na caixa de diálogo Selecionar Casos.

Isso abre a caixa de diálogo Selecionar Casos, na qual é possível selecionar um intervalo de números de observações (linhas).

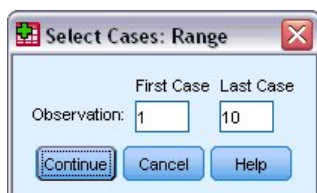


Figura 87. Caixa de diálogo Selecionar Intervalo de Casos

- **Primeiro Caso.** Insira os valores iniciais de data e/ou hora para o intervalo. Se nenhuma variável de data for definida, insira o número de observação inicial (número da linha no Editor de Dados, a menos que Dividir Arquivo esteja ligado). Se você não especificar um valor de Último Caso, todos os casos a partir da data/hora inicial até o final da série de tempo serão selecionados.
- **Último Caso.** Insira os valores finais de data e/ou hora para o intervalo. Se nenhuma variável de data for definida, insira o número de observação final (número da linha no Editor de Dados, a menos que Dividir Arquivo esteja ligado). Se você não especificar um valor de Primeiro Caso, todos os casos a partir do início da série de tempo até a data/hora final serão selecionados.

Para os dados da série de tempo com variáveis de data definidas, você pode selecionar um intervalo de datas e/ou horas com base nas variáveis de data definidas. Cada caso representa observações em uma hora diferente e o arquivo é classificado em ordem cronológica.



Figura 88. Caixa de diálogo Selecionar Intervalo de Casos (série de tempo)

Para gerar variáveis de data para dados da série de tempo:

2. Nos menus, escolha:

**Dados > Definir Datas...**

## Tratamento de casos não selecionados

É possível escolher uma das alternativas a seguir para o tratamento de casos não selecionados:

- **Filtrar casos não selecionados.** Os casos não selecionados não são incluídos na análise, mas permanecem no conjunto de dados. É possível usar os casos não selecionados mais tarde na sessão se você desligar a filtragem. Se você selecionar uma amostra aleatória ou se selecionar casos com base em uma expressão condicional, isso gerará uma variável denominada *filter\_\$* com um valor de 1 para casos selecionados e um valor de 0 para casos não selecionados.
- **Copiar casos selecionados para um novo conjunto de dados.** Os casos selecionados são copiados para um novo conjunto de dados, deixando o conjunto de dados original não afetado. Os casos não



selecionados não são incluídos no novo conjunto de dados e são deixados em seu estado original no conjunto de dados original.

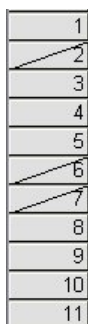
- **Excluir casos não selecionados.** Os casos não selecionados são excluídos do conjunto de dados. Os casos excluídos podem ser recuperados apenas saindo do arquivo sem salvar nenhuma mudança e, em seguida, reabrindo o arquivo. A exclusão de casos será permanente se você salvar as mudanças no arquivo de dados.

*Nota:* Se você excluir os casos não selecionados e salvar o arquivo, os casos não poderão ser recuperados.

## Status de seleção de caso

---

Se você tiver selecionado um subconjunto de casos, mas não tiver descartado os casos não selecionados, os casos não selecionados serão marcados no Editor de Dados com uma linha diagonal por meio do número da linha.



1
2
3
4
5
6
7
8
9
10
11

*Figura 89. Status de seleção de caso*



---

## Capítulo 10. 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.** Ao trabalhar em uma sintomatologia padronizada de comportamento anoréxico/bulímico, os pesquisadores <sup>1</sup> fizeram um estudo de 55 adolescentes com distúrbios 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 <sup>2</sup>, foi solicitado a 52 estudantes que classificassem as combinações de 15 situações e 15 comportamentos em uma escala de 10 pontos que vão de 0="extremamente apropriado" a 9="extremamente inapropriado". Na média de indivíduos, os valores são assumidos como dissimilaridades.

---

<sup>1</sup> 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.

<sup>2</sup> 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 <sup>3</sup>, foi solicitado a 21 estudantes de MBA da Wharton School e a seus cônjuges que classificassem 15 itens de café da manhã em ordem de preferência com 1="mais preferido" a 15="menos preferido". 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 <sup>4</sup> diz respeito a reclamações de danos para 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 <sup>5</sup>, uma empresa interessada em fazer propaganda de um novo limpador de carpete quer examinar a influência de cinco fatores na preferência do consumidor — design do pacote, nome da marca, preço, um selo *Good Housekeeping* e uma garantia de reembolso. 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* contêm 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.

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

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

<sup>5</sup> 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 escoragem 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 diz respeito a imagens recebidas de seis marcas de café gelado <sup>6</sup>. 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,

<sup>6</sup> 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.

provincia, 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 de Stillman" <sup>7</sup>. 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 máquina <sup>8</sup> 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.** Bell <sup>9</sup> apresentou uma tabela para ilustrar possíveis grupos sociais. Guttman <sup>10</sup> usou uma parte desta tabela, na qual cinco variáveis que descrevem coisas, como interação social, sentimentos de pertencer a um grupo, proximidade física de 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 apresentações em sala de aula), público (por exemplo, públicos de jornal ou televisão), multidões (como uma multidão mas com interação muito mais intensa), grupos primários (íntimos), grupos secundários (voluntários) e a comunidade moderna (confederação livre resultante de grande proximidade física e de uma 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

---

<sup>7</sup> 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.

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

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

<sup>10</sup> 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 juizes treinados (mais um entusiasta) a 300 apresentações de ginástica. Cada linha representa uma apresentação separada; os juizes visualizaram as mesmas apresentações.
- **kinship\_dat.sav.** Rosenberg e Kim <sup>11</sup> se dispuseram a 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/](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. Os pesquisadores anteriores <sup>12, 13</sup>, entre outros, encontraram não linearidades entre essas variáveis, que dificultaram 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.
- **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

---

<sup>11</sup> 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.

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

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

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\_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 escoragem foi registrada. Cada caso



no arquivo de dados representa um estagiário separado e registra o grupo ao qual ele foi designado e a escoragem 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, porcas e tachas <sup>14</sup>.
- **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 <sup>15</sup> que diz respeito a danos em navios cargueiros causados por 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 depois de ter sido limpo usando procedimentos na opção Preparação de Dados.
- **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.
- **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.

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

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

- **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 censurados por intervalo e foi apresentado e analisado em outro lugar <sup>16</sup>.
- **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 <sup>17</sup>.
- **verd1985.sav.** Este arquivo de dados diz respeito a uma pesquisa de opinião <sup>18</sup>. 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.
- **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 das crianças <sup>19</sup>. 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.

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

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

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

- **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.

---

<sup>19</sup> 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:

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

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

## Marcas comerciais

---

IBM, o logotipo IBM e [ibm.com](http://ibm.com) são marcas comerciais ou marcas registradas da International Business Machines Corp., registradas em 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](http://www.ibm.com/legal/copytrade.shtml).

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

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

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

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

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

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





# Índice Remissivo

## A

Access (Microsoft) [11](#)  
arquivos de amostra  
  posição [79](#)  
arquivos de banco de dados  
  lendo [11](#)  
arquivos de dados de texto  
  lendo [14](#)  
arquivos de planilha  
  lendo [8](#)  
  lendo nomes de variáveis [8](#)  
arquivos de sintaxe  
  abrindo [61](#)  
Arquivos do Excel  
  lendo [8](#)  
Assistente de Base de Dados [11](#)  
Assistente de data e hora [67](#)  
Assistente de Importação de Texto [14](#)

## C

calculando novas variáveis [65](#)  
casos  
  ordenando [71, 73](#)  
  seleção [73](#)  
classificando casos [71](#)  
colando sintaxe  
  de uma caixa de diálogo [59](#)  
contagens  
  tabelas de contagens [25](#)  
criar rótulos de variáveis [21](#)

## D

dados categóricos  
  medidas de resumo [25](#)  
dados contínuos [25](#)  
dados de escala [25](#)  
dados de intervalo [25](#)  
dados de sequência  
  inserindo dados [20](#)  
dados nominais [25](#)  
dados numéricos [19](#)  
dados ordinais [25](#)  
dados proporcionais [25](#)  
dados qualitativos [25](#)  
dados quantitativos [25](#)

## E

editando tabelas dinâmicas [40](#)  
Editor de Dados  
  inserindo dados não numéricos [20](#)  
  inserindo dados numéricos [19](#)  
entrada de dados [19, 20](#)

estratos  
  criando em tabelas dinâmicas [40](#)  
Excel (Microsoft)  
  exportando resultados para [50](#)  
  exportando os resultados  
    HTML [57](#)  
  exportando resultados  
    para Excel [50](#)  
    para PowerPoint [50](#)  
    para Word [50](#)  
expressões condicionais [67](#)

## F

ferramenta de Ajuda de Sintaxe [60](#)  
funções em expressões [66](#)

## G

gráficos  
  barra [26, 31](#)  
  criando gráficos [31](#)  
  histogramas [28](#)  
gráficos de barras [26](#)

## H

histogramas [28](#)  
HTML  
  exportando os resultados [57](#)

## I

inserindo dados  
  não numéricos [20](#)  
  numéricos [19](#)

## J

janelas de sintaxe  
  codificação de cores [60](#)  
  colando comandos [59](#)  
  conclusão automática [60](#)  
  editando comandos [60](#)  
  executando comandos [59, 61](#)  
  pontos de interrupção [61](#)

## M

medidas de resumo  
  dados categóricos [25](#)  
  variáveis de escala [27](#)  
movendo  
  elementos em tabelas dinâmicas [39](#)  
  itens no Visualizador [37](#)

## N

nível de medição [25](#)

## O

ocultando linhas e colunas em tabelas dinâmicas [41](#)

## P

PowerPoint (Microsoft)

exportando resultados para [50](#)

processamento de divisão de arquivo [71](#)

## R

recodificando valores [63](#)

rótulos de valor

controlando a exibição no Visualizador [22](#)

designando [22](#)

variáveis numéricas [22](#)

rótulos de variáveis

criação [21](#)

## S

selecionando casos [73](#)

sintaxe [59](#)

subconjuntos de casos

amostra aleatória [75](#)

com base em datas e horas [76](#)

excluindo casos não selecionados [76](#)

expressões condicionais [74](#)

filtrando casos não selecionados [76](#)

se a condição for satisfeita [74](#)

seleção [73](#)

## T

tabelas de frequências [25](#)

tabelas dinâmicas

acessando definições [38](#)

bandejas de dinamização [39](#)

editando [40](#)

estratos [40](#)

formatando [40](#)

formatos de célula [42](#)

ocultando linhas e colunas [41](#)

ocultando pontos decimais [42](#)

tipos de dados da célula [42](#)

transpondo linhas e colunas [39](#)

tipos de dados

para variáveis [22](#)

transpondo (invertendo) linhas e colunas em tabelas

dinâmicas [39](#)

## V

valores ausentes

omissos do sistema [23](#)

para variáveis não numéricas [24](#)

para variáveis numéricas [23](#)

valores omissos do sistema [23](#)

variáveis

rótulos [21](#)

tipos de dados [22](#)

variáveis de data e hora [67](#)

variáveis de escala

medidas de resumo [27](#)

Visualizador

movendo saída [37](#)

ocultando e mostrando saída [37](#)

## W

Word (Microsoft)

exportando resultados para [50](#)



