

Aula 2:

‘Que dados tenho?’

Estatística descritiva univariada
(percentagens, médias, medianas e outras
medidas de estatística descritiva)

Docente: Daniela Craveiro
dcraveiro@iseg.ulisboa.pt

**No final desta aula,
@s alun@s deverão:**

- Escolher as estatísticas descritivas mais adequadas em função da natureza da variável
- Compreender o impacto de dados extremos e omissos na distribuição de dados
- Saber produzir tabelas de frequências, tabelas com medidas de tendência central e de dispersão
- Saber identificar e codificar valores omissos e extremos
- Saber reportar estatísticas descritivas no contexto de um relatório de dados

A estatística descritiva serve para resumir e descrever os nossos dados
Como descrever a variável que nos interessa?

Podemos recorrer a

- Distribuição das frequências (e outros indicadores genéricos)
- Medidas de tendência central (sumarizam a distribuição da variável)
- Medidas de dispersão (indicam o grau de variância da distribuição)

- **Distribuição das Frequências**

	DEFINIÇÃO	TIPO DE VARIÁVEL	OBSERVAÇÕES
Frequência	Número observações com um determinado valor na variável	Todos	-
Proporção	Número observações com um determinado valor na variável, dividido pelo número total de observações.	Todos	• Expressa numa escala de 0-1
Percentagem	Número observações com um determinado valor na variável, dividido pelo número total de observações, multiplicado por 100.	Todos	• Expressa numa escala de 0-100
Rácio	Número de observações com um valor, dividido pelo número de observações com outro valor	<u>Nominal</u> Ordinal	• Expressa em valores ancorados (ex: por cada 100 indivíduos na categoria X, existem N indivíduos na categoria Y)

INDICADORES GENÉRICOS

Frequências, rácios, taxas

Frequências

Frequência absoluta

$$F_i = F_1 + F_2 + \dots + F_n$$

Frequência relativa / proporção.

$$f_i = (F_1 + F_2 + \dots + F_n) / n$$

máximo=1

Frequência relativa / percentagem.

$$f_i = (F_1 + F_2 + \dots + F_n) / n * 100$$

máximo=100

Rácios

Rácio = $n \text{ categoria A} / n \text{ categoria B}$

Taxa

Taxa = $\text{valor} / \text{capacidade} * 100$

Taxa de variação

Taxa = $(\text{valor final} - \text{valor inicial}) / \text{inicial} * 100$

Taxa de variação média

Taxa var média = $(\text{valor final} - \text{valor inicial})^{1/p} - 1 * 100$

- **Medidas de tendência central**

	DEFINIÇÃO	TIPO DE VARIÁVEL	OBSERVAÇÕES
Moda (ou <i>valor modal</i>)	Valor mais frequente da variável	Todos	• Adequada se a distribuição da variável é enviesada • Não é sensível a valores extremos
Média	Valor médio das observações na variável	Intervalar, Razão	• Adequada se distribuição da variável não é enviesada • É sensível a valores extremos • Quando os dados estão agrupados (ex: grupos de idade), calcula-se com base no valor mediano dos intervalos
Mediana	Valor na distribuição que divide a variável em duas partes iguais (50% para cima e 50% para baixo)	Ordinal, Intervalar, Razão	• Adequada se a distribuição da variável é enviesada • Não é sensível a valores extremos

MEDIDAS DE LOCALIZAÇÃO

Moda, Mediana, Média

Moda

A moda corresponde à categoria ou valor com maior frequência. É facilmente identificada quando os dados são resumidos numa forma tabular.

Mediana

É o valor central (ou média dos dois valores centrais) da distribuição ordenada dos valores.



Média

A média aritmética obtém-se com a soma dos valores da variável, dividida pelo nº de casos observados.

Qual é a tendência central?



ANÁLISE DE DADOS EM GRH

• Medidas de dispersão

	DEFINIÇÃO	TIPO DE VARIÁVEL	OBSERVAÇÕES
Intervalo de Variação	Diferença entre o valor mais alto e o valor mais baixo da variável	Intervalar, Razão	• É sensível a valores extremos
Amplitude Interquartis	Diferença entre o valor do 3º Quartil e o valor do 1º Quartil	Intervalar, Razão	• Não é sensível a valores extremos
Rácio Interdecis	90/10: Valor do 9º decil dividido pelo valor do 1º decil	Razão	• Alternativa: 80/20 • Não é sensível a valores extremos
Variância	Média das diferenças dos valores observados em relação à média, elevadas ao quadrado	Intervalar, Razão	• = 0: Não há variância • Quanto maior, maior variância • É sensível a valores extremos • Difícil de interpretar
Desvio-Padrão	Raiz quadrada da variância.	Intervalar, Razão	• Mais fácil de interpretar • Usa a mesma escala de valores da variável • É sensível a valores extremos

MEDIDAS DE DISPERSÃO

Amplitude, Variância, Desvio padrão

Amplitude

Traduz a diferença entre os dois valores extremos da variável. Também referido por Intervalo de variação ou Amplitude do Intervalo de variação:

$$I_v = \text{Max} - \text{Min}$$

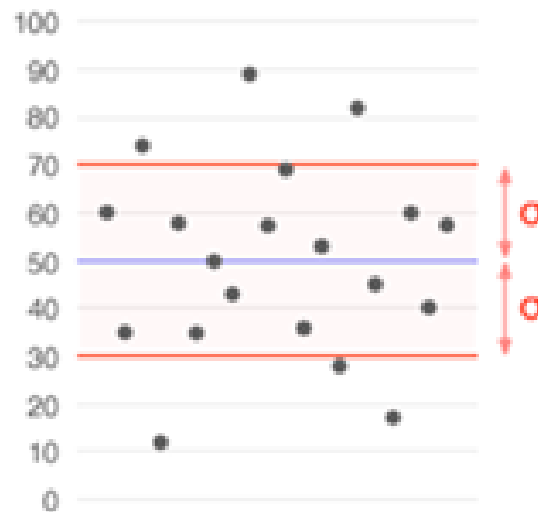
Variância

Indica-nos o quão próximos os valores estão da média. Resulta da soma do quadrado dos desvios em relação à média, a dividir pelo total de observações.

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Desvio padrão

Raiz quadrada da variância. Expressa os desvios na escala da variável.



- **O que são percentis, decis e quartis?**
 - São estatísticas descritivas que são calculadas ordenando a variável em função do número de observações, e depois dividindo-a em partes iguais
 - Tomando como exemplo uma variável que mede os salários na economia
 - **Percentil** – Apura-se dividindo a variável em 100 partes iguais.
 - 1º Percentil - ex: 1% das pessoas têm salários até €100
 - 99º Percentil - ex: 99% das pessoas têm salários até €5000

- **O que são percentis, decis e quartis?**
 - São estatísticas descritivas que são calculadas ordenando a variável em função do número de observações, e depois dividindo-a em partes iguais
 - Tomando como exemplo uma variável que mede os salários na economia
 - **Decil** – Apura-se dividindo a variável em **10** partes iguais.
 - 1º Decil - ex: 10% das pessoas têm salários até €400
 - 9º Decil - ex: 90% das pessoas têm salários até €2000

- **O que são percentis, decis e quartis?**
 - São estatísticas descritivas que são calculadas ordenando a variável em função do número de observações, e depois dividindo-a em partes iguais
 - Tomando como exemplo uma variável que mede os salários na economia
 - **Quartil** – Apura-se dividindo a variável em 4 partes iguais.
 - 1º Quartil - ex: 25% das pessoas têm salários até €600
 - 3º Quartil - ex: 75% das pessoas têm salários até €1500

Análise descritiva univariada

Distribuição das Frequências

Objetivo: Gerar tabela de frequências para a variável “department”

Distribuição de Frequências

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Frequências'
- Selecionar variável 'department'
- Colocar na caixa de 'Variáveis'

A

The screenshot shows the top menu bar of IBM SPSS Statistics. The 'Analisar' (Analyze) menu is open, and the path 'Estatística Descritiva' (Descriptive Statistics) > 'Frequências...' (Frequencies...) is highlighted. A red circle and the letter 'A' mark this path.

B

C

The screenshot shows the 'Frequências' (Frequencies) dialog box. In the 'Variável(is):' list on the left, the variable 'department' is selected and highlighted. A red circle and the letter 'B' mark this selection. A red circle and the letter 'C' mark the arrow button used to move the selected variable to the 'Variável(is):' list.

Visualização de dados **Visualização de variável**

O processador do IBM SPSS Statistics está pronto | Unicode: ON

Distribuição de Frequências

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Frequências'
- Selecionar variável 'department'
- Colocar na caixa de 'Variáveis'
- Selecionar 'OK'

A

The screenshot shows the IBM SPSS Statistics menu bar. The 'Analisar' (Analyze) menu is open, and the path 'Estatística Descritiva' (Descriptive Statistics) > 'Frequências...' (Frequencies...) is highlighted. A red circle and the letter 'A' mark this path.

B

C

D

The screenshot shows the 'Frequências' (Frequencies) dialog box. The variable 'department' is selected in the 'Variáveis:' list. A red circle and the letter 'B' mark the list. A red circle and the letter 'C' mark the arrow button used to move the selected variable to the 'Variável(is):' box. A red circle and the letter 'D' mark the 'OK' button at the bottom of the dialog box.

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode: ON

Tabela de Frequências

- O resultado é publicado no 'Visualizador de Resultados'

The screenshot shows the IBM SPSS Statistics Output Viewer window. The left pane displays a tree view of the output, with 'Frequências' selected under the 'Log' folder. The main pane displays the following content:

FREQUENCIES VARIABLES=department
/ORDER=ANALYSIS.

Frequências

Estatísticas

department

N	Válido	Omisso
	5000	0

department

	Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido				
Production	2129	42.6	42.6	42.6
Logistics	938	18.8	18.8	61.3
Sales	494	9.9	9.9	71.2
Admin	404	8.1	8.1	79.3
Financial	282	5.6	5.6	84.9
Marketing	250	5.0	5.0	89.9
IT	228	4.6	4.6	94.5
HR	180	3.6	3.6	98.1
Audit	95	1.9	1.9	100.0
Total	5000	100.0	100.0	

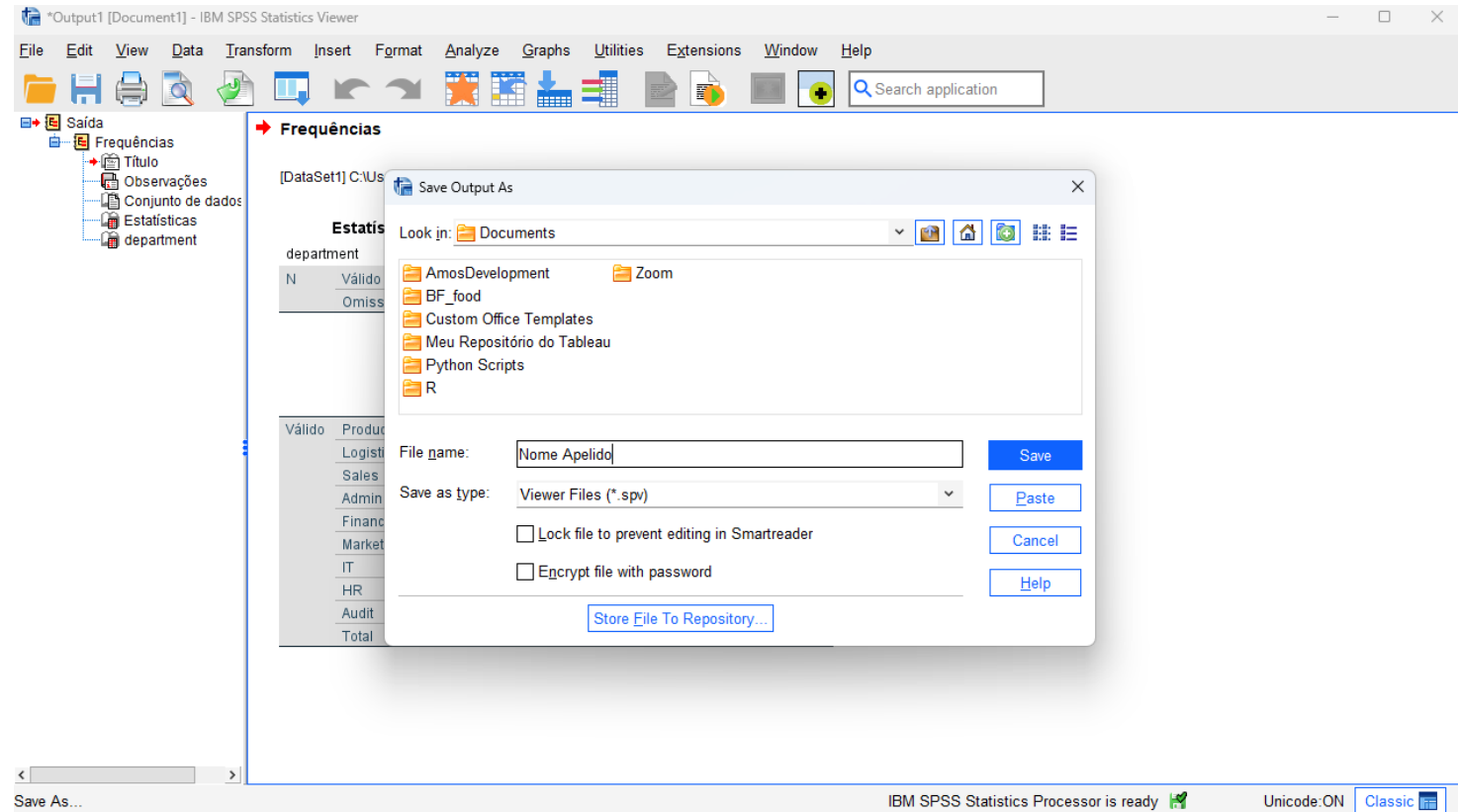
Imprimir

O processador do IBM SPSS Statistics está pronto

Unicode: ON

Tabela de Frequências

- Esta é altura para guardar o **Output (Nome Apelido)** para ser carregado na Pasta Teams! (avaliação contínua)



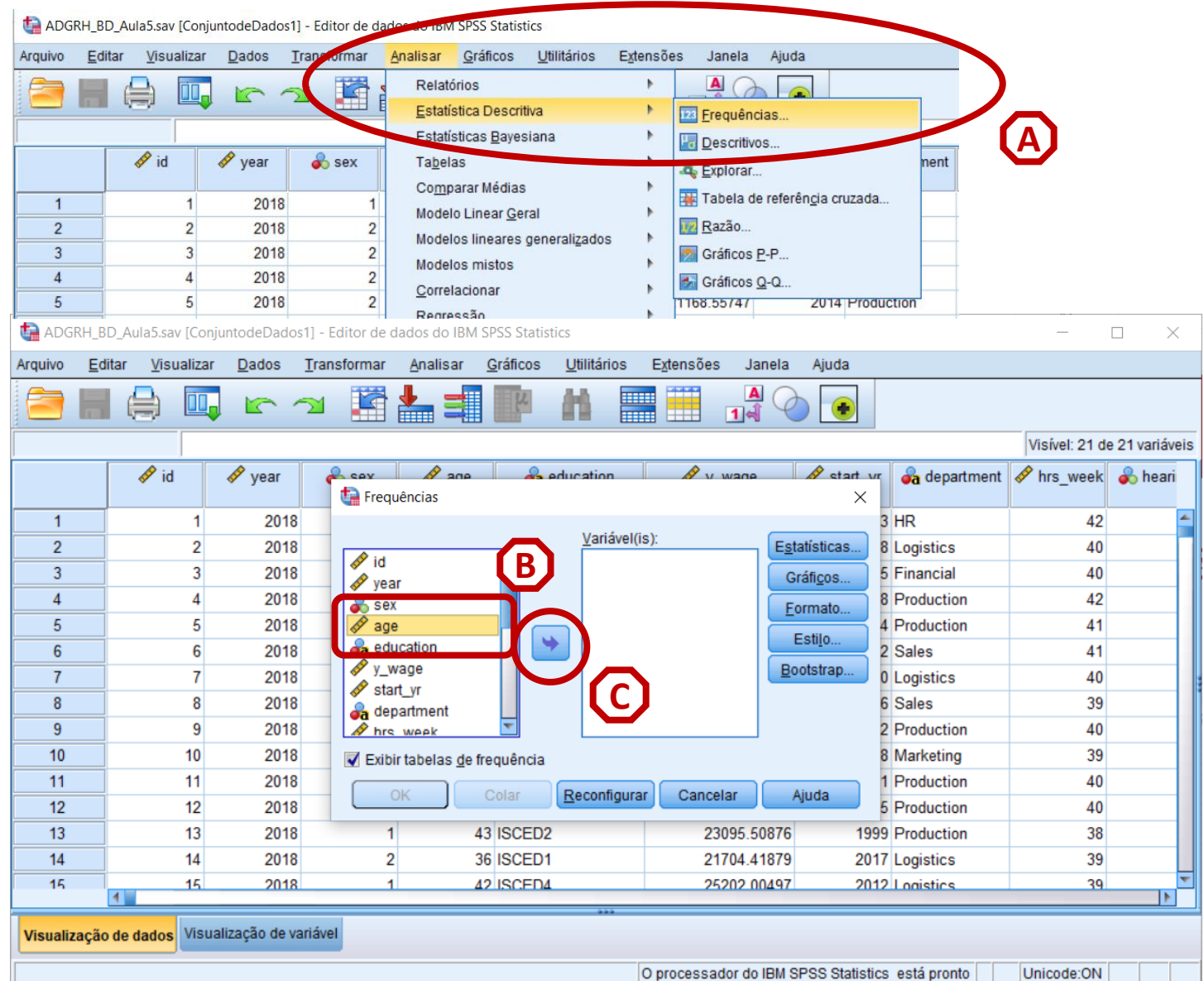
Análise descritiva univariada

Medidas de Tendências Central

Objetivo: Gerar estatísticas descritivas para a variável “age”

Medidas de Tendência Central

- Abrir base de dados
- Selecionar 'Analisar' /
'Estatística Descritiva' /
'Frequências' A
- Selecionar variável 'age' B
- Colocar na caixa de 'Variáveis' C



Medidas de Tendência Central

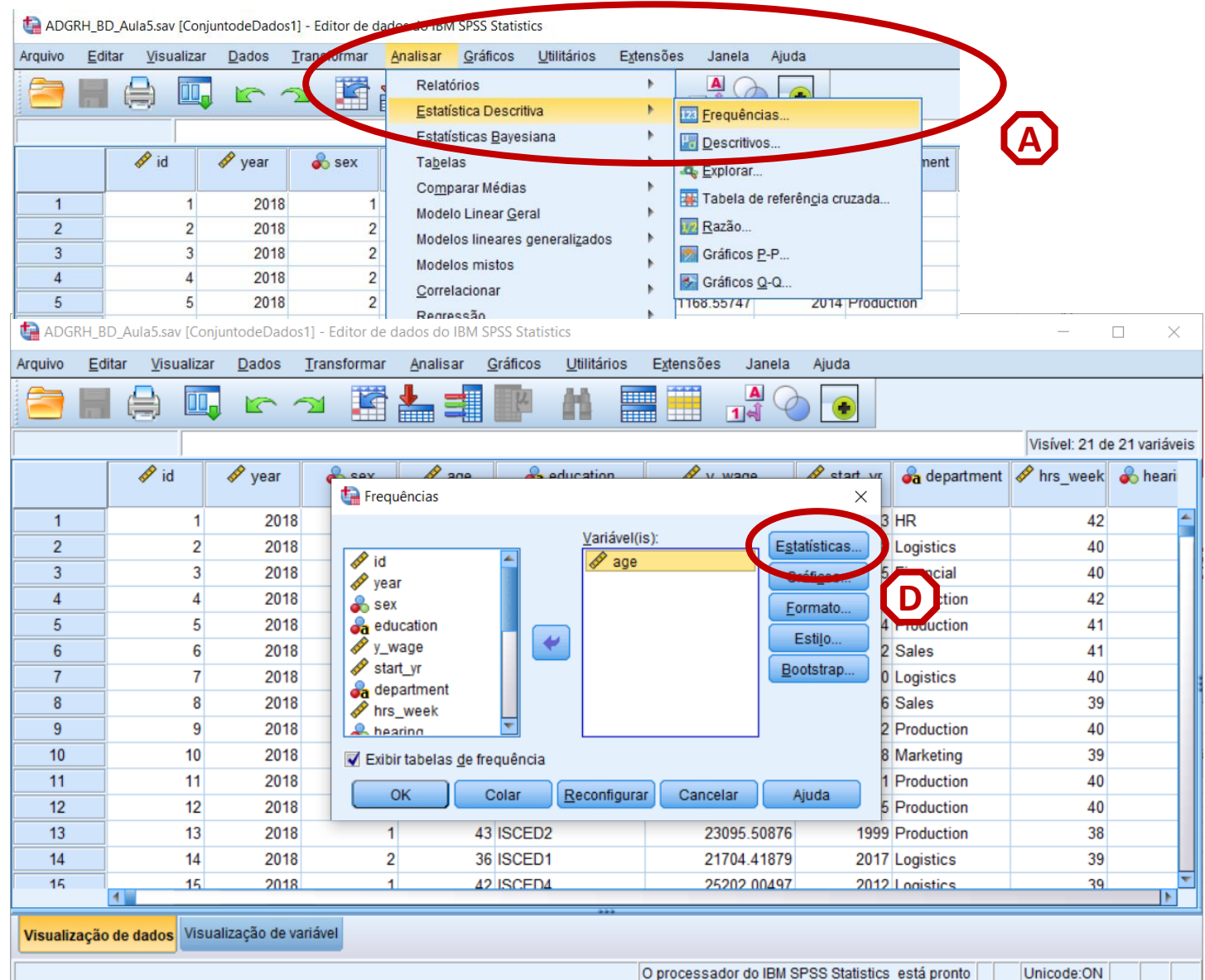
- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Frequências'
- Selecionar variável 'age'
- Colocar na caixa de 'Variáveis'
- Selecionar 'Estatísticas'

A

B

C

D



Medidas de Tendência Central

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Frequências'
- Selecionar variável 'age'
- Colocar na caixa de 'Variáveis'
- Selecionar 'Estatísticas'
- Selecionar Medidas de Tendência Central
- Selecionar 'Continuar' / 'OK'

A

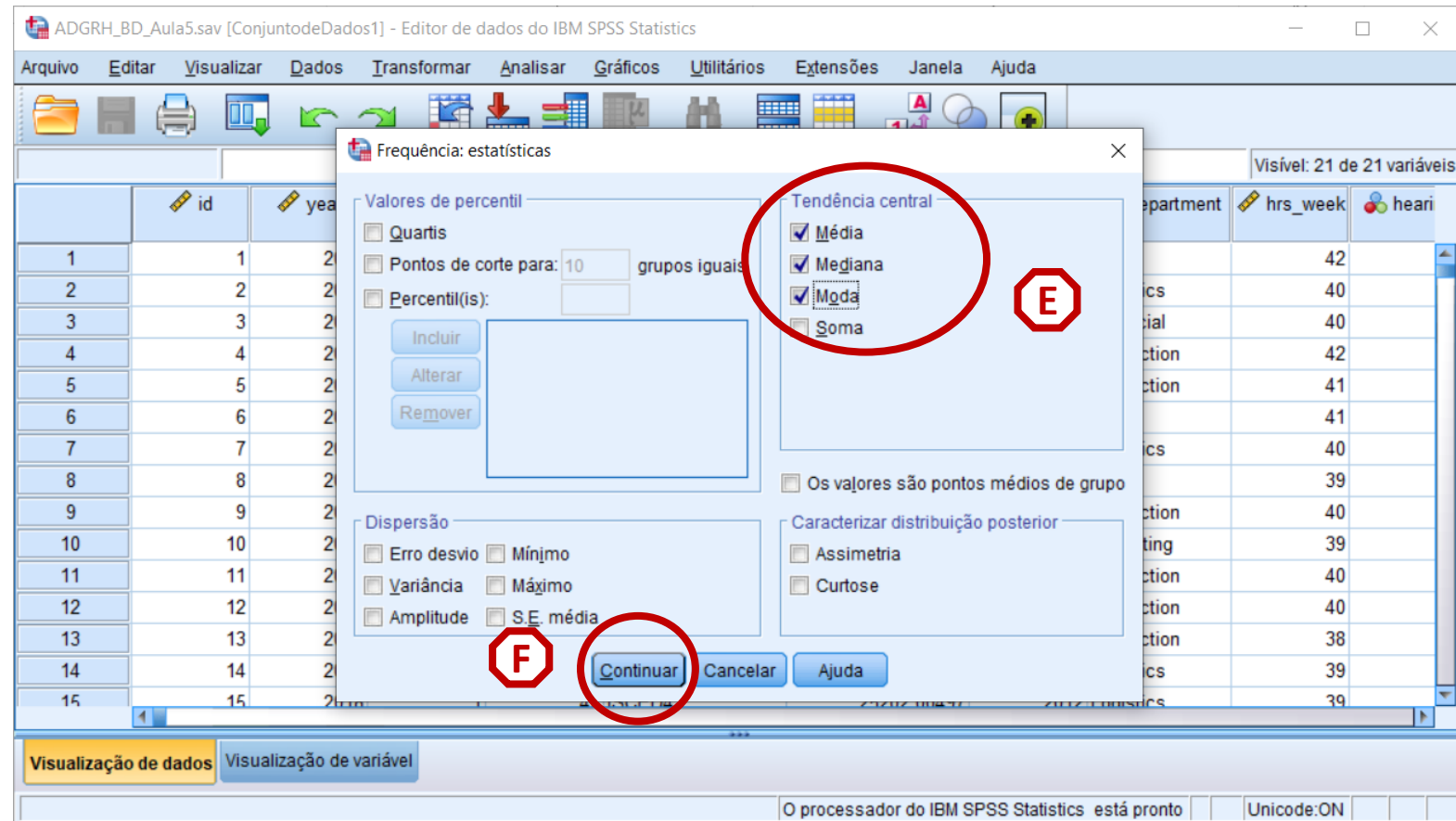
B

C

D

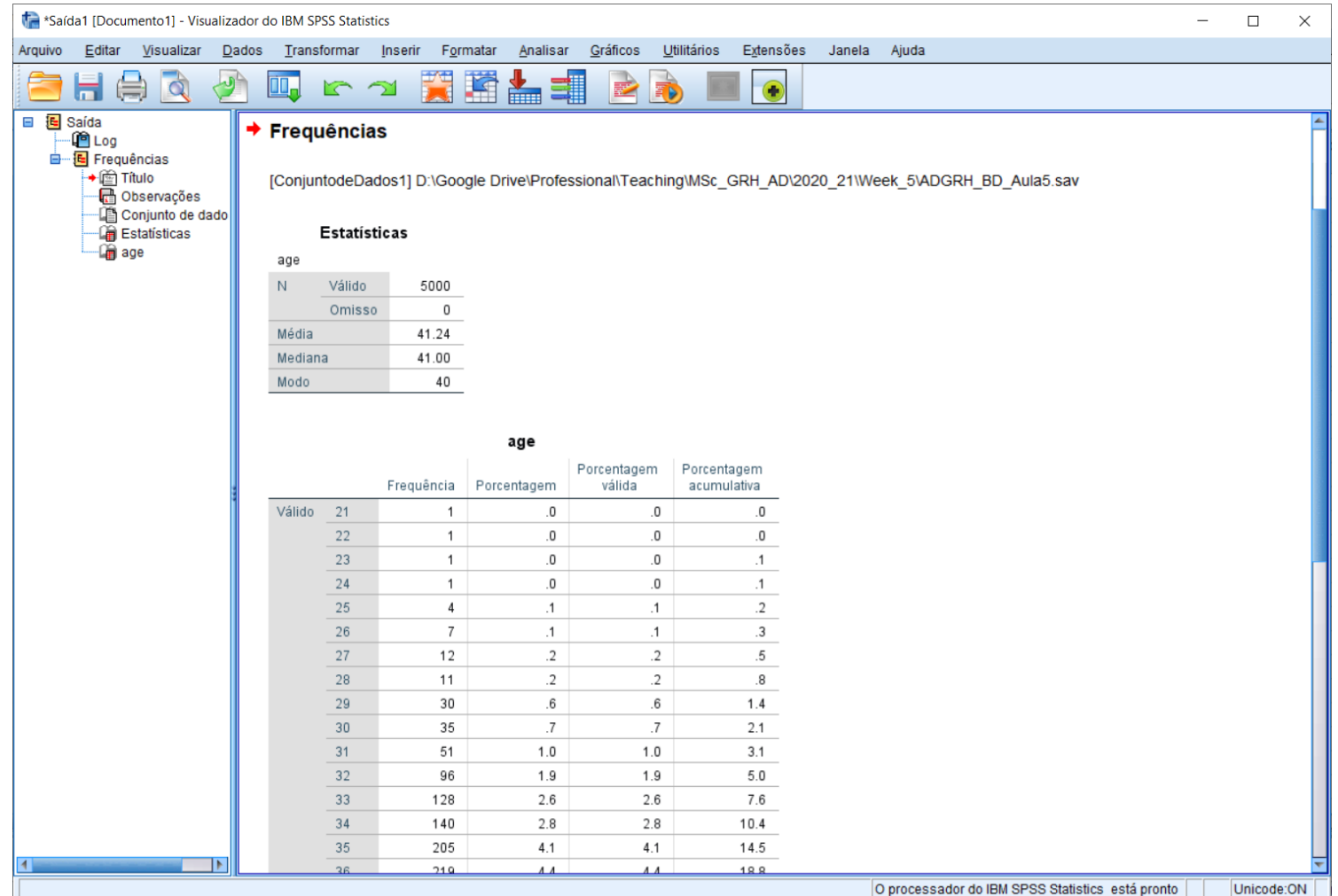
E

F



Medidas de Tendência Central

- O resultado é publicado no 'Visualizador de Resultados'



The screenshot shows the IBM SPSS Statistics 'Visualizador de Resultados' (Results Viewer) window. The left pane displays a tree structure with 'Saída' (Output) expanded, showing 'Log', 'Frequências' (Frequencies), 'Título' (Title), 'Observações' (Observations), 'Conjunto de dados' (Data Set), 'Estatísticas' (Statistics), and 'age'. The right pane shows the 'Frequências' output for the variable 'age' from the file 'D:\Google Drive\Professional\Teaching\MSc_GRH_AD\2020_21\Week_5\ADGRH_BD_Aula5.sav'. It includes a summary table of statistics and a detailed frequency table.

Frequências
[ConjuntodeDados1] D:\Google Drive\Professional\Teaching\MSc_GRH_AD\2020_21\Week_5\ADGRH_BD_Aula5.sav

Estatísticas

age		
N	Válido	Omisso
	5000	0
Média	41.24	
Mediana	41.00	
Modo	40	

age

		Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido	21	1	.0	.0	.0
	22	1	.0	.0	.0
	23	1	.0	.0	.1
	24	1	.0	.0	.1
	25	4	.1	.1	.2
	26	7	.1	.1	.3
	27	12	.2	.2	.5
	28	11	.2	.2	.8
	29	30	.6	.6	1.4
	30	35	.7	.7	2.1
	31	51	1.0	1.0	3.1
	32	96	1.9	1.9	5.0
	33	128	2.6	2.6	7.6
	34	140	2.8	2.8	10.4
	35	205	4.1	4.1	14.5
	36	219	4.4	4.4	18.9

Análise descritiva univariada

Medidas de Dispersão

Objetivo: Gerar estatísticas descritivas para a variável “evaluation”

Medidas de Dispersão

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Frequências'
- Selecionar variável 'evaluation'
- Colocar na caixa de variáveis
- Selecionar 'Estatísticas'

The image displays two screenshots of the IBM SPSS Statistics software interface, illustrating the steps to generate a frequency table for a specific variable.

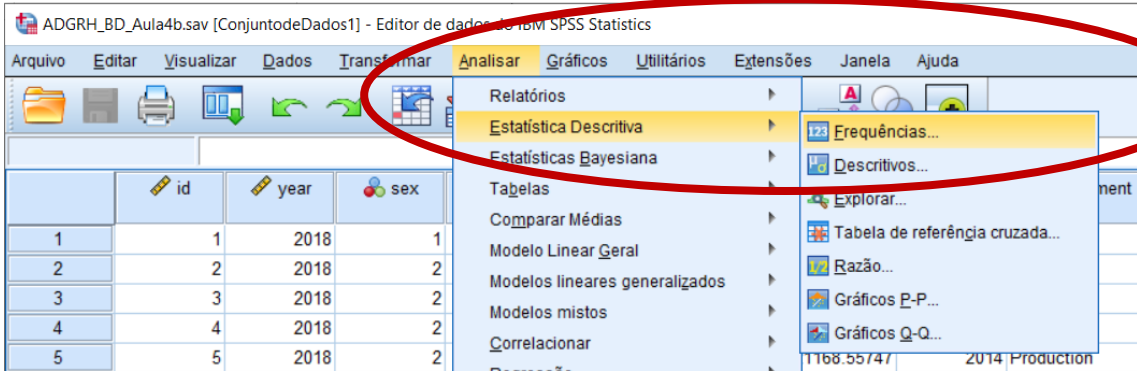
Screenshot 1 (Top): The 'Analisar' (Analyze) menu is open, showing the path: **Analisar** > **Estatística Descritiva** (Descriptive Statistics) > **Frequências...** (Frequencies...). A red circle highlights the 'Frequências...' option, and a red octagon labeled 'A' is placed next to it.

Screenshot 2 (Bottom): The 'Frequências' (Frequencies) dialog box is open. The variable 'evaluation' is selected in the list on the left and moved to the 'Variável(is):' (Variable(s)) box on the right. A red circle highlights the 'evaluation' variable, and a red octagon labeled 'B' is placed next to it. In the 'Variável(is):' box, the 'Estatísticas...' (Statistics...) button is highlighted with a red circle, and a red octagon labeled 'D' is placed next to it. A red octagon labeled 'C' is placed over the arrow button used to move the variable from the list to the box. The 'Exibir tabelas de frequência' (Display frequency tables) checkbox is checked. The 'OK' button is visible at the bottom of the dialog box.

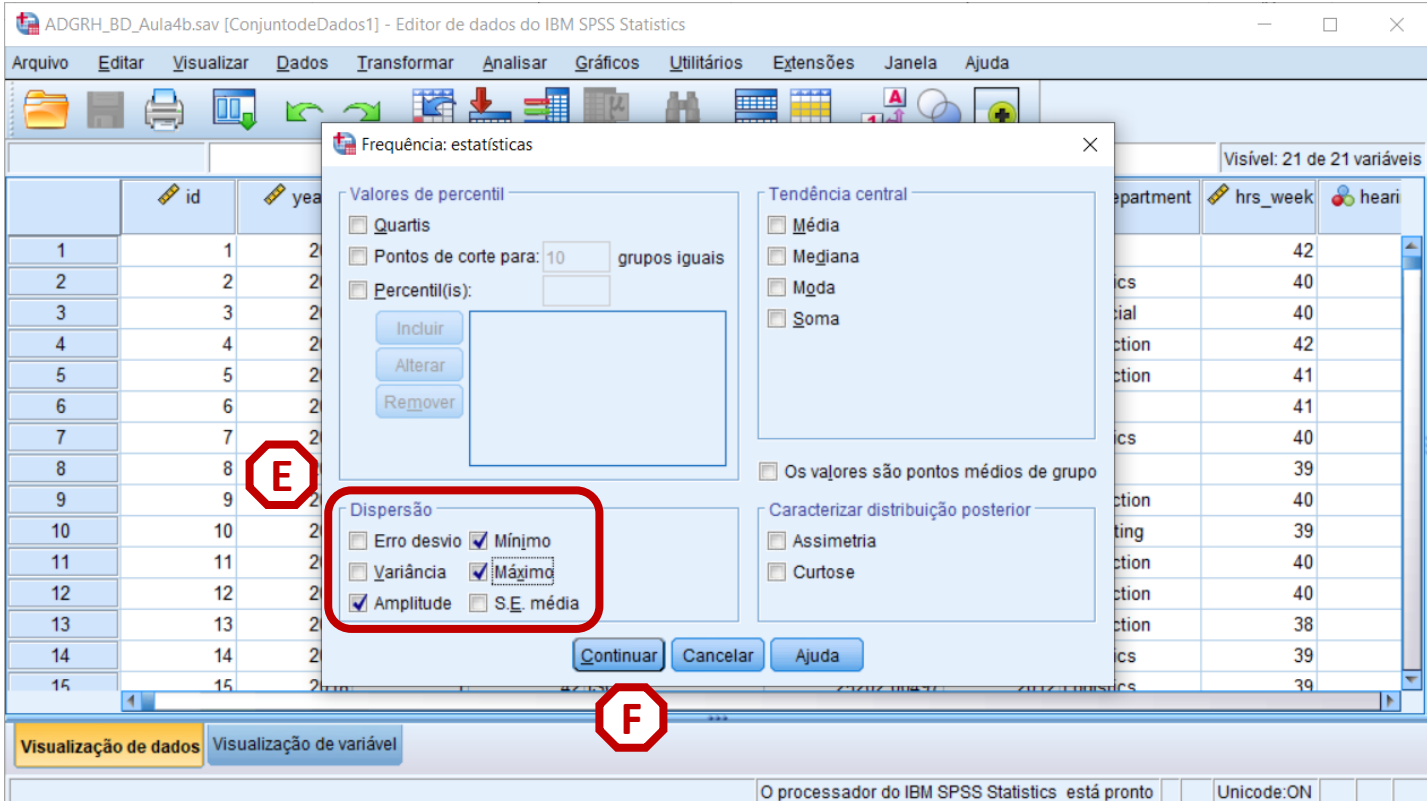
Medidas de Dispersão

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Frequências'
- Selecionar variável 'evaluation'
- Colocar na caixa de variáveis
- Selecionar 'Estatísticas'
- Selecionar as Medidas de Dispersão baseadas em valores absolutos (Min, Max, Amplitude)
- Selecionar 'Continuar' / 'OK'

A



B



C

D

E

F

Medidas de Dispersão

- O resultado é publicado no 'Visualizador de Resultados'

The screenshot shows the 'Visualizador do IBM SPSS Statistics' window. The left sidebar displays a tree structure of the output, with 'Frequências' selected under the 'evaluation' variable. The main area shows the command: `FREQUENCIES VARIABLES=evaluation /STATISTICS=RANGE MINIMUM MAXIMUM /ORDER=ANALYSIS.` Below this, the 'Estatísticas' table is displayed, followed by the 'evaluation' frequency table.

FREQUENCIES VARIABLES=evaluation
/STATISTICS=RANGE MINIMUM MAXIMUM
/ORDER=ANALYSIS.

→ **Frequências**

Estatísticas

evaluation		
N	Válido	Omissão
	5000	0
Intervalo	19.73	
Mínimo	41.85	
Máximo	61.58	

evaluation

	Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido	41.85	1	.0	.0
	43.66	1	.0	.0
	44.05	1	.0	.1
	44.11	1	.0	.1
	44.16	1	.0	.1
	44.40	1	.0	.1
	44.49	1	.0	.1
	44.62	1	.0	.2
	44.68	1	.0	.2
	44.93	1	.0	.2
	45.05	1	.0	.2
	45.08	1	.0	.2
	45.16	1	.0	.3

Medidas de Dispersão

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Frequências'
- Selecionar variável 'evaluation'
- Colocar na caixa de variáveis
- Selecionar 'Estatísticas'

The image displays two screenshots of the IBM SPSS Statistics software interface, illustrating the steps to generate a frequency table for a specific variable.

Screenshot 1 (Top): The 'Analisar' (Analyze) menu is open, showing the path: **Analisar** > **Estatística Descritiva** (Descriptive Statistics) > **Frequências...** (Frequencies...). A red circle highlights the 'Frequências...' option, and a red octagon labeled 'A' is placed next to it.

Screenshot 2 (Bottom): The 'Frequências' (Frequencies) dialog box is open. The variable 'evaluation' is selected in the list on the left and moved to the 'Variável(is):' (Variable(s)) box on the right. A red circle highlights the 'evaluation' variable, and a red octagon labeled 'B' is placed next to it. In the 'Variável(is):' box, the 'Estatísticas...' (Statistics...) button is highlighted with a red circle, and a red octagon labeled 'D' is placed next to it. A red octagon labeled 'C' is placed over the arrow button used to move the variable from the list to the box. The 'Exibir tabelas de frequência' (Display frequency tables) checkbox is checked. The background shows a data table with columns: id, year, sex, age, education, v. wage, start_yr, department, hrs_week, and heari.

Medidas de Dispersão

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Frequências'
- Selecionar variável 'evaluation'
- Colocar na caixa de variáveis
- Selecionar 'Estatísticas'
- Selecionar 'Quartis'

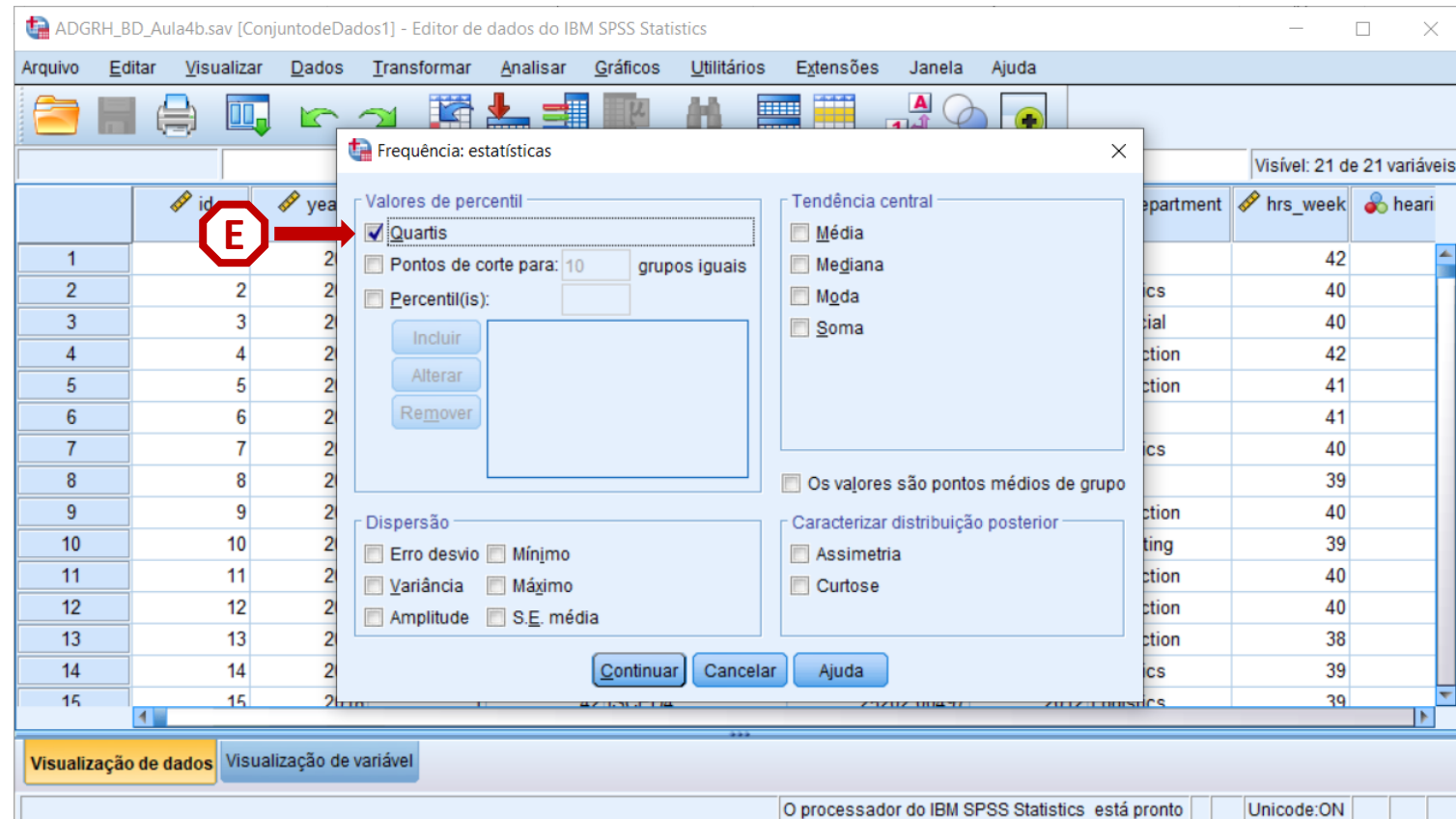
A

B

C

D

E



Medidas de Dispersão

- Selecionar 'Percentis'
- Inserir 10º Percentil (= 1º Decil)
- Selecionar 'Incluir'

Exercício: Inserir 9º Decil

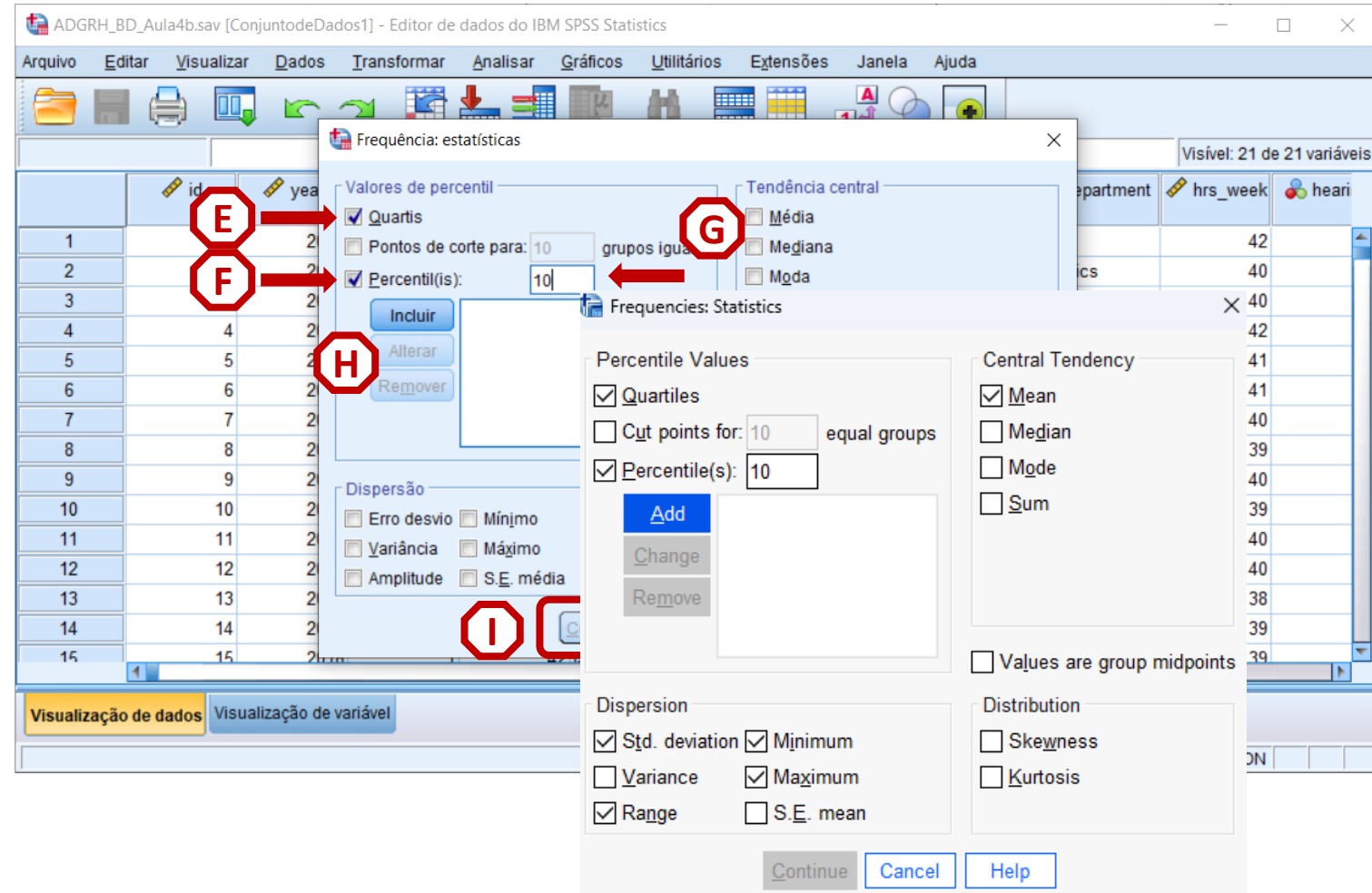
- Selecionar 'Continuar' / 'OK'

F

G

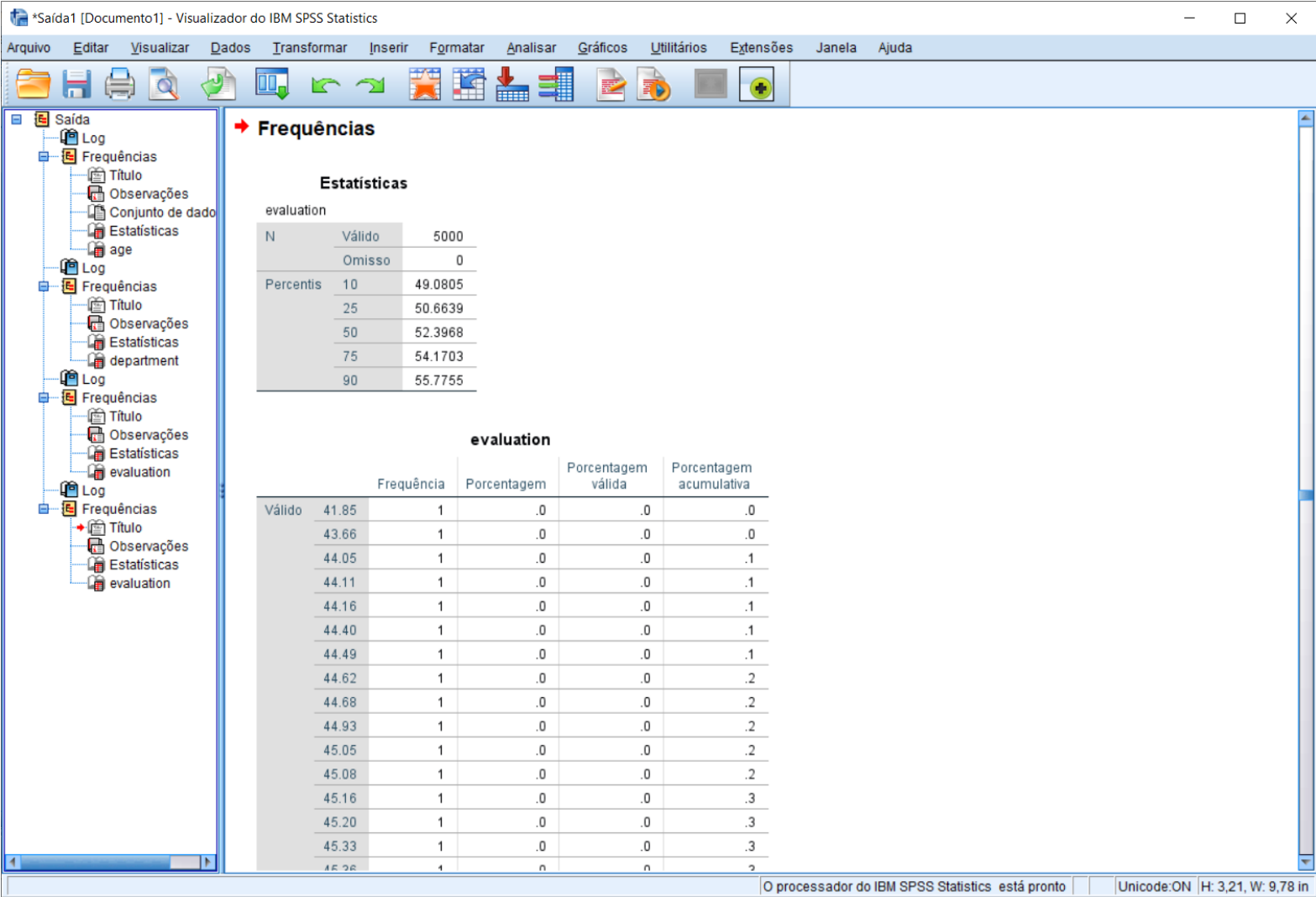
H

I



Medidas de Dispersão

- O resultado é publicado no ‘Visualizador de Resultados’



- Qual a melhor forma de visualizar o grau de dispersão da variável?

Gráfico de Caixa

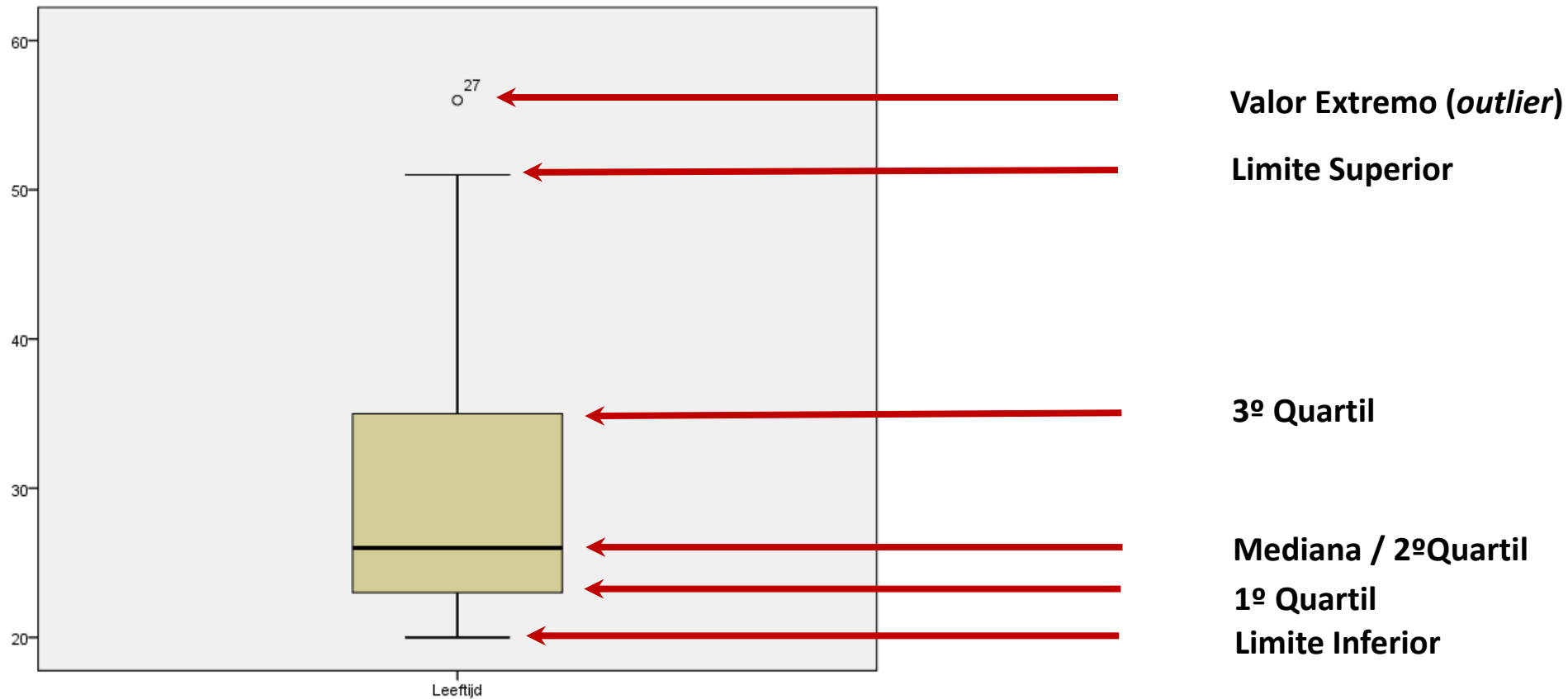


Gráfico de Caixa

- Selecionar 'Gráficos' / 'Construtor de Gráfico'

A

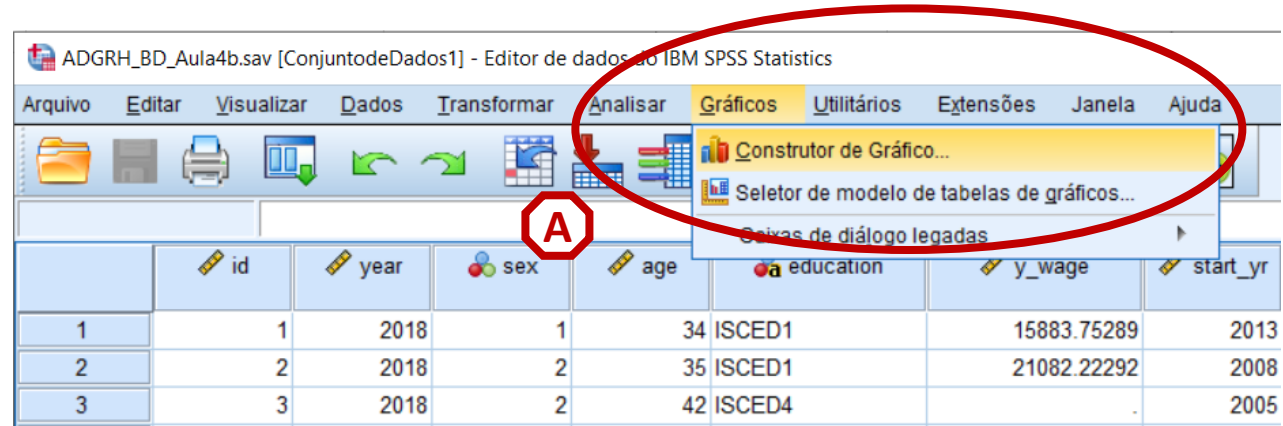


Gráfico de Caixa

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'Boxplot'
- Duplo-clique no Gráfico de Caixa (simples)

A

B

C

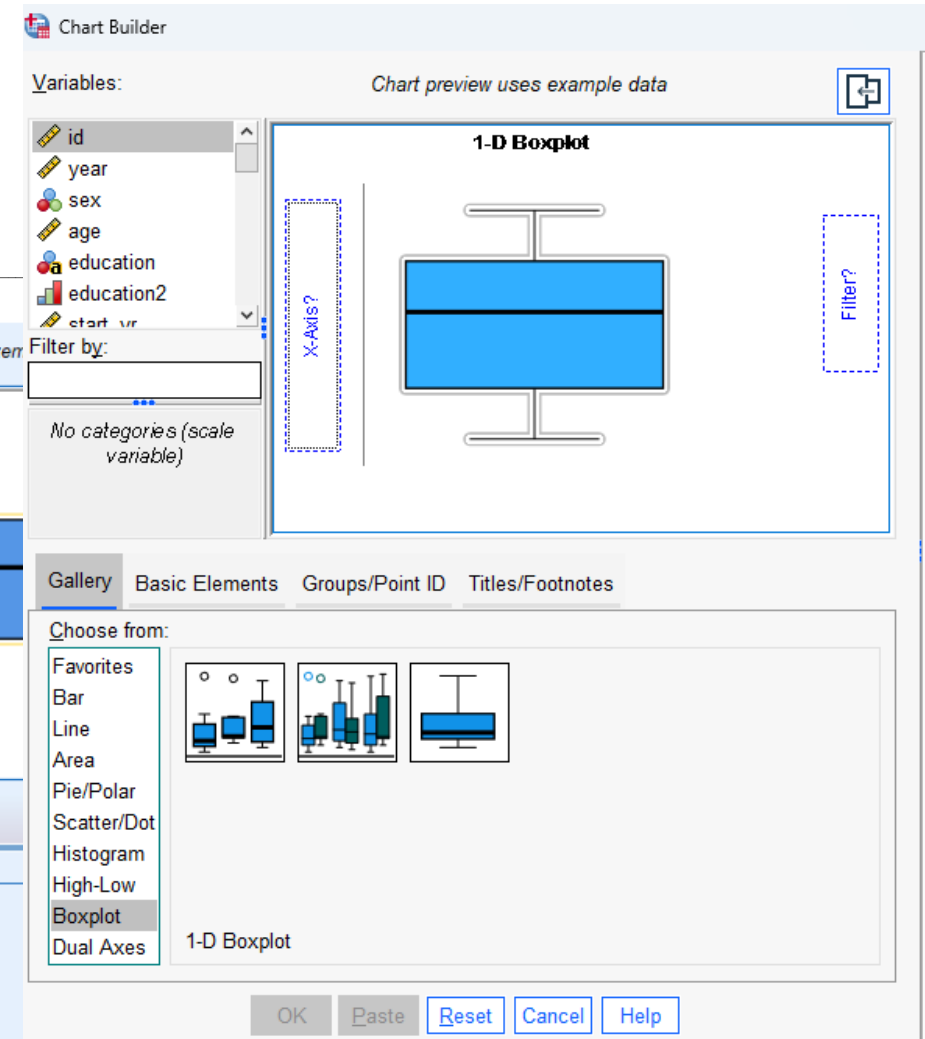
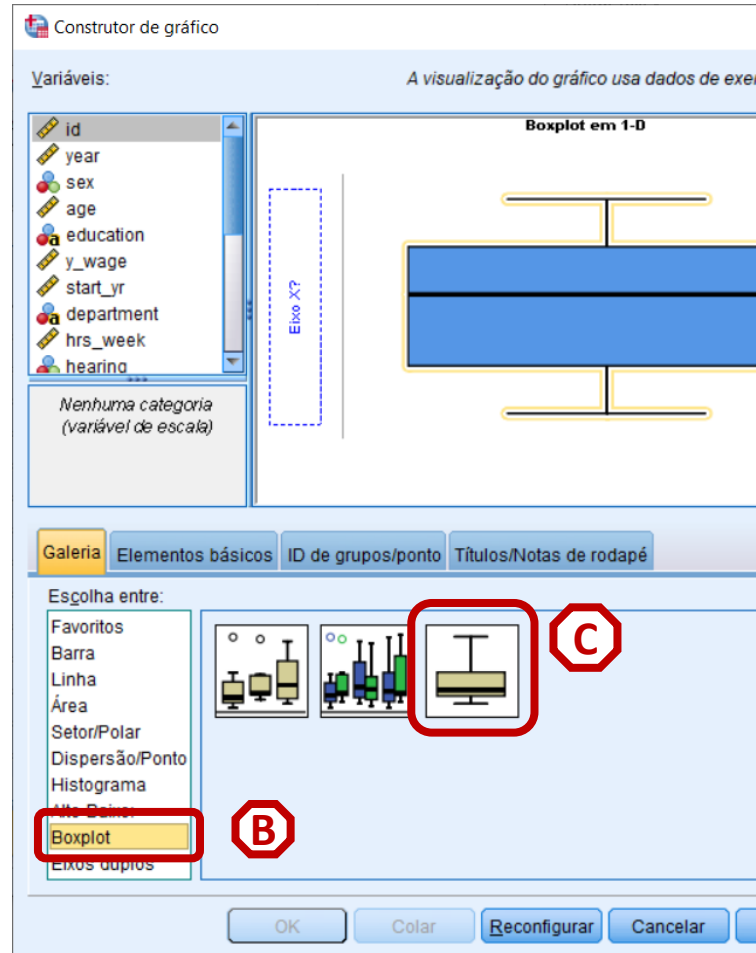


Gráfico de Caixa

- Selecionar ‘Gráficos’ / ‘Construtor de Gráfico’
- Selecionar ‘Boxplot’
- Duplo-clique no Gráfico de Caixa (simples)
- Seleccionar a variável ‘evaluation’
- E arrastar para o ‘Eixo X’

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- back
- muscle_upper
- muscle_lower
- headache
- injury
- anxiety
- fatigue
- absent_nr
- workcond_sat
- evaluation

uma categoria (vel de escala)

Boxplot em 1-D

Eixo X?

Filtro?

Galeria Elementos básicos ID de grupos/ponto Títulos/Notas de rodapé

Escolha entre:

- Favoritos
- Barra
- Linha
- Área
- Setor/Polar
- Dispersão/Ponto
- Histograma
- Alto-Baixo:
- Boxplot
- Eixos duplos

OK Colar Reconfigurar Cancelar Ajuda

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de:

Caixa1

X-Eixo1 (Caixa1)

Título 1

Estatísticas

Variável:

Estatística:

Boxplot

Configurar parâmetros...

Exibir barra de erros

Representação de Barras de Erros

- Intervalos de confiança
- Nível (%): 95
- Erro padrão
- Multiplicador: 2
- Desvio padrão
- Multiplicador: 2

Gráfico de Caixa

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'Boxplot'
- Duplo-clique no Gráfico de Caixa (simples)
- Seleccionar a variável 'evaluation'
- E arrastar para o 'Eixo X'
- Selecionar 'OK'

A

B

C

D

E

F

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- back
- muscle_upper
- muscle_lower
- headache
- injury
- anxiety
- fatigue
- absent_nr
- workcond_sat
- evaluation

uma categoria (nível de escala)

Boxplot em 1-D de evaluation

Filtro?

Galeria Elementos básicos ID de grupos/ponto Títulos/Notas de rodapé

Escolha entre:

- Favoritos
- Barra
- Linha
- Área
- Setor/Polar
- Dispersão/Ponto
- Histograma
- Alto-Baixo:
- Boxplot
- Eixos duplos

OK Colar Reconfigurar Cancelar Ajuda

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de:

Caixa1

X-Eixo1 (Caixa1)

Título 1

Estatísticas

Variável:

Estatística:

Boxplot

Configurar parâmetros...

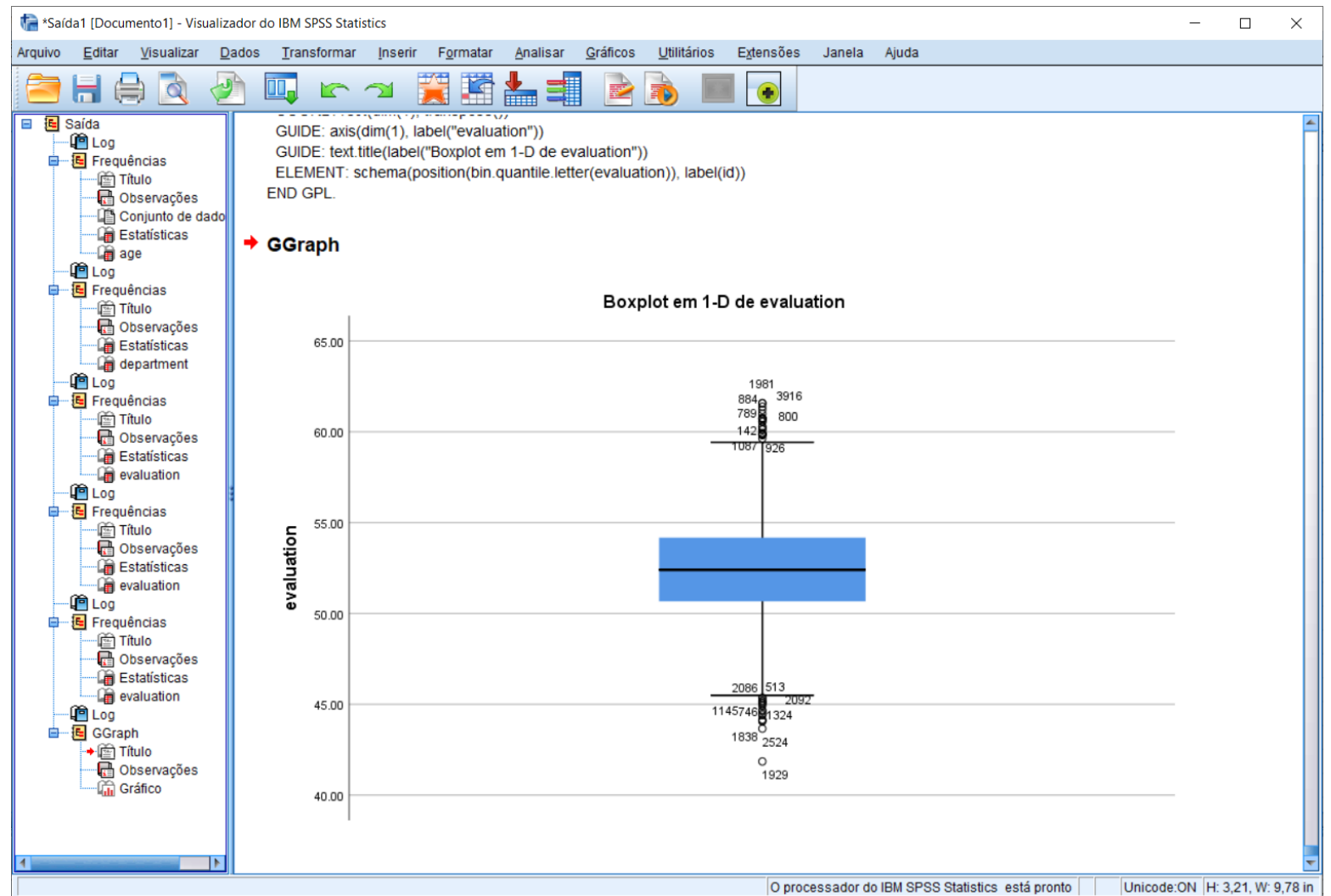
☒ Exibir barra de erros

Representação de Barras de Erros

- ☒ Intervalos de confiança
- Nível (%): 95
- ☒ Erro padrão
- Multiplicador: 2
- ☒ Desvio padrão
- Multiplicador: 2

Gráfico de Caixa

- O gráfico é publicado no 'Visualizador de Resultados'



Com a exploração de dados muitas vezes identificamos imprecisões e omissões dos nossos dados que temos de dar conta

Como limpar a base de dados?

Tipicamente essa fase envolve:

- Identificar/Recodificar 'Valores Omissos'
- Identificar/Recodificar 'Valores Extremos'

ANÁLISE DE DADOS EM GRH

- Por vezes, iremos encontrar valores omissos na nossa base de dados ...

ADGRH_BD_Aula3b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Visível: 15 de 15 variáveis

	id	year	gender	age	marital	labforce	empstat	wage	hearing	eye	memory
1	126962	2015	1	52,00	4	2	4	99999999,00	1	1	1
2	16717	2015	1	32,00	1	3	1	6141,00	1	1	1
3	22360	2015	2	46,00	1	3	1	46000,00	1	1	1
4	23780	2015	1	44,00	1	1	4	120000,00	.	.	.
5	134027	2015	1	12,00	6	1	.	99999999,00	.	.	.
6	13215	2015	1	21,00	6	2	4	6500,00	1	1	1
7	173046	2015	2	52,00	6	2	3	99999999,00	1	1	1
8	195945	2015	1	8,00	6	1	.	99999999,00	.	.	.
9	97355	2015	2	18,00	6	3	1	7002,00	1	1	1
10	180306	2015	2	15,00	6	3	2	99999999,00	1	1	1
11	170018	2015	1	67,00	4	3	1	99999999,00	2	1	1
12	159824	2015	1	66,00	5	2	5	99999999,00	1	1	1
13	67213	2015	2	,00	6	1	.	99999999,00	.	.	.
14	100988	2015	1	29,00	4	3	1	20700,00	1	1	1
15	18096	2015	2	14,00	6	1	.	99999999,00	.	.	.
16	187745	2015	2	40,00	4	2	1	10000,00	1	1	1

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF

O que pode explicar a existência de valores omissos?

- **Inquéritos:**
 - ‘Não Sabe’/‘Não Responde’: Particularmente comum em questões delicadas (ex. rendimentos);
 - ‘Não Aplicável’: Depende de como o pesquisador o define.
- **Estatísticas Oficiais:**
 - Dados não estão disponíveis.

Por que é que a existência de dados omissos é um problema?

- Reduz o número de observações
- Pode enviesar os resultados da nossa análise

Em geral os valores omissos aparecem de três formas:

- System Missing: ‘.’
- Sob a forma de texto: ‘Não Aplicável’
- Como um valor: ‘99’; ‘99999’

Limpar a Base de Dados

Identificar/Recodificar ‘Valores Omissos’

Definir como Valor Omisso

Recodificar como Valor Omisso

Substituir o Valor Omisso por Outro Valor (ex: Média)

Limpar a Base de Dados

Identificar/Recodificar ‘Valores Omissos’

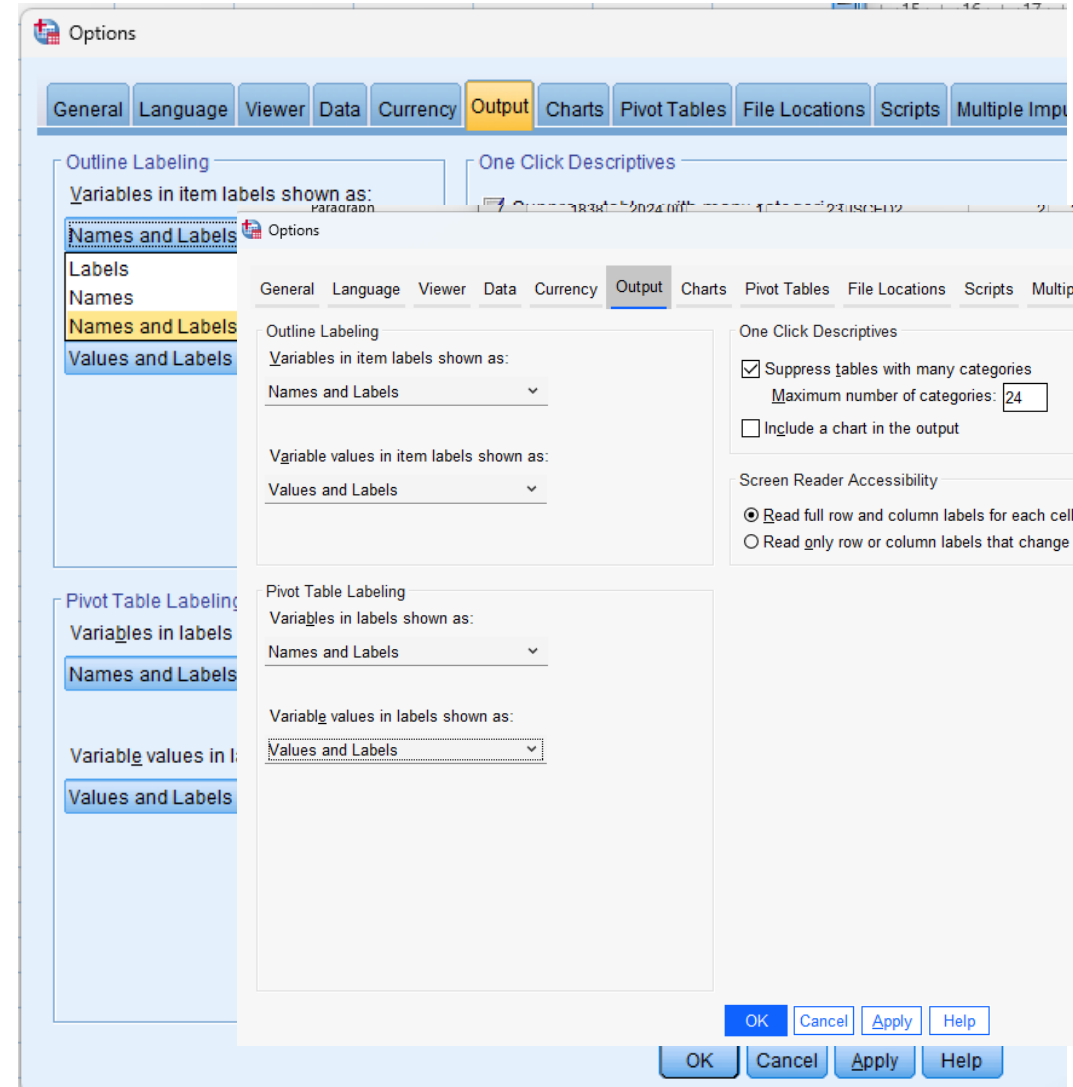
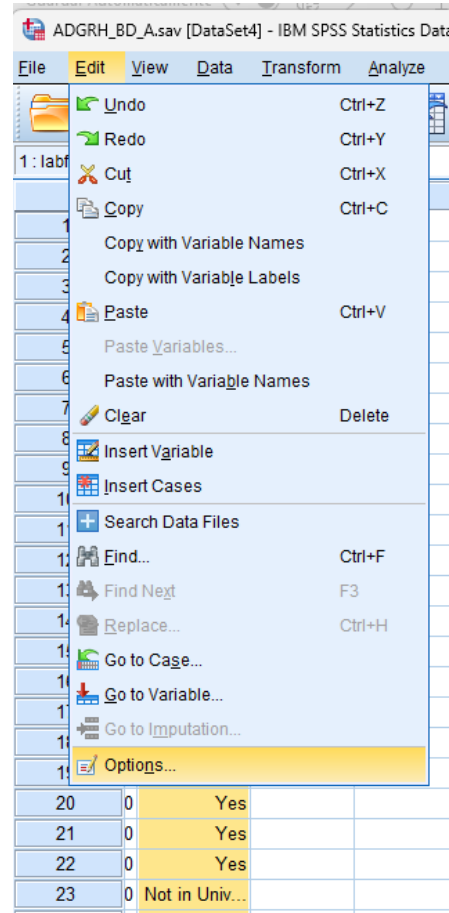
Definir como Valor Omisso

Recodificar como Valor Omisso

Substituir o Valor Omisso por Outro Valor (ex: Média)

Importante!
Definir as opções da
forma como o código
surge.

Edit/Options
Output
“Names and Labels”
- nas três subsecções



Limpar a Base de Dados

Identificar ‘Valores Omissos’

Objetivo: Definir categoria ‘Not in Universe’ da variável labforce como valor omissos

Definir Valores Omissos

- A variável 'labforce' está codificada da seguinte forma:
 - 1: 'Not in Universe'
 - 2: 'No'
 - 3: 'Yes'
- 'Not in Universe' é tratada como se fosse uma categoria normal. **A**
- Objectivo
 - Definir 'Not in Universe' como valor omissos.

The screenshot shows the IBM SPSS Statistics interface. The 'Frequencies' dialog box is open, with 'id' and 'year' selected in the 'Variable(s):' list. The 'Display Variable Names' option is selected. The 'Statistics' button is highlighted. The output window shows the 'Estadísticas' table for 'In the Labour Force'.

Estadísticas

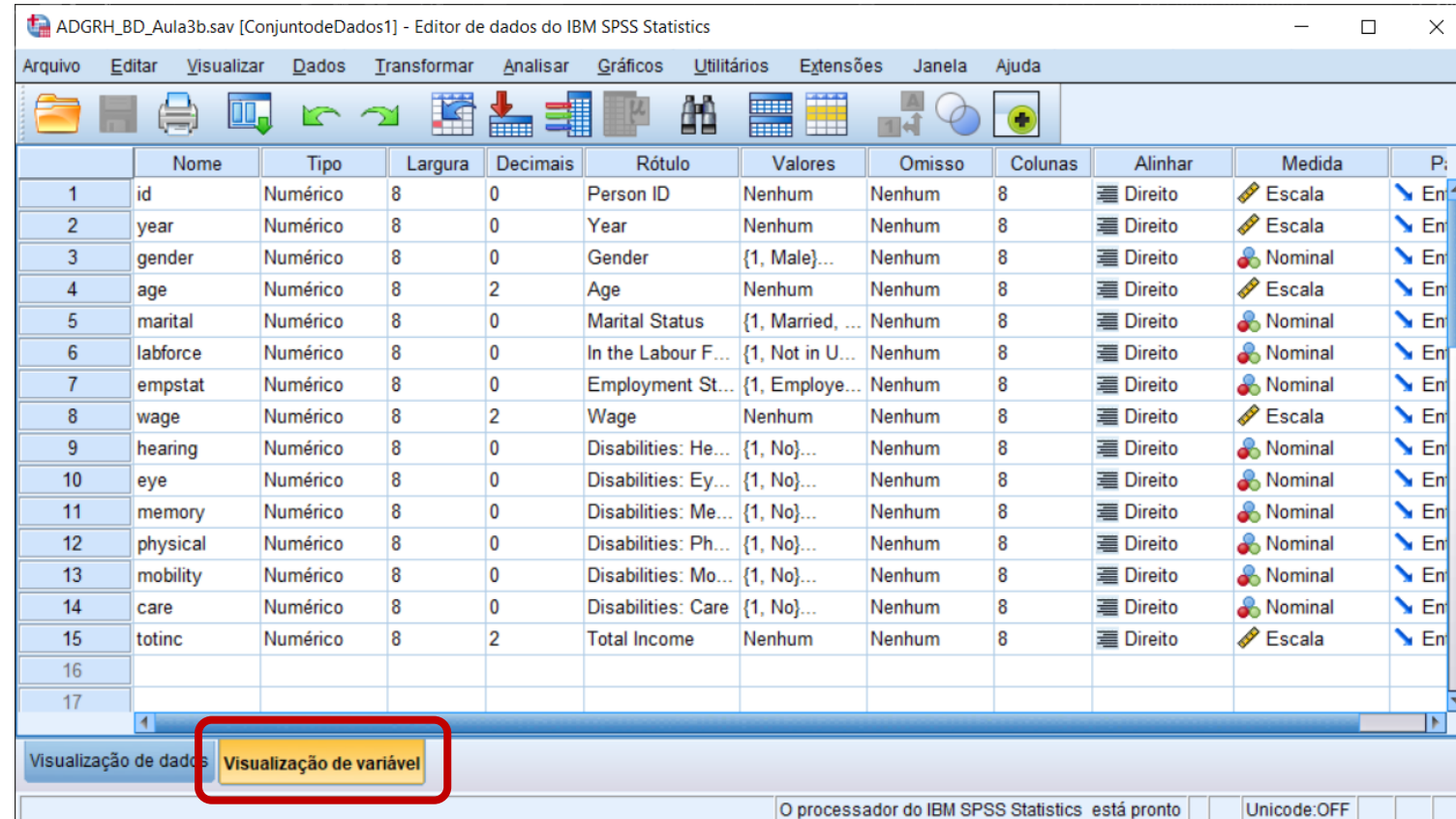
In the Labour Force	
N	Válido 5000
	Omisso 0

In the Labour Force

	Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido				
Not in Universe	1178	23,6	23,6	23,6
No	1489	29,8	29,8	53,3
Yes	2333	46,7	46,7	100,0
Total	5000	100,0	100,0	

Definir Valores Omissos

- O SPSS oferece-nos uma maneira muito simples de fazer isso...
- Selecionar 'Visualização de variável'



Definir Valores Omissos

- O SPSS oferece-nos uma maneira muito simples de fazer isso...
- Selecionar 'Visualização de variável'
- Na coluna 'Omisso', clicar na célula da variável 'labforce'

A

B

ADGRH_BD_Aula3b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

	Nome	Tipo	Largura	Decimais	Rótulo	Valores	Omisso	Colunas	Alinhar	Medida	P:
1	id	Numérico	8	0	Person ID	Nenhum	Nenhum	8	Direito	Escala	En
2	year	Numérico	8	0	Year	Nenhum	Nenhum	8	Direito	Escala	En
3	gender	Numérico	8	0	Gender	{1, Male}...	Nenhum	8	Direito	Nominal	En
4	age	Numérico	8	2	Age	Nenhum	Nenhum	8	Direito	Escala	En
5	marital	Numérico	8	0	Marital Status	{1, Married, ...}	Nenhum	8	Direito	Nominal	En
6	labforce	Numérico	8	0	In the Labour F...	{1, Not in U...	Nenhum	8	Direito	Nominal	En
7	empstat	Numérico	8	0	Employment St...	{1, Employee...	Nenhum	8	Direito	Nominal	En
8	wage	Numérico	8	2	Wage	Nenhum	Nenhum	8	Direito	Escala	En
9	hearing	Numérico	8	0	Disabilities: He...	{1, No}...	Nenhum	8	Direito	Nominal	En
10	eye	Numérico	8	0	Disabilities: Ey...	{1, No}...	Nenhum	8	Direito	Nominal	En
11	memory	Numérico	8	0	Disabilities: Me...	{1, No}...	Nenhum	8	Direito	Nominal	En
12	physical	Numérico	8	0	Disabilities: Ph...	{1, No}...	Nenhum	8	Direito	Nominal	En
13	mobility	Numérico	8	0	Disabilities: Mo...	{1, No}...	Nenhum	8	Direito	Nominal	En
14	care	Numérico	8	0	Disabilities: Care	{1, No}...	Nenhum	8	Direito	Nominal	En
15	totinc	Numérico	8	2	Total Income	Nenhum	Nenhum	8	Direito	Escala	En
16											
17											

Visualização de dados Visualização de variável

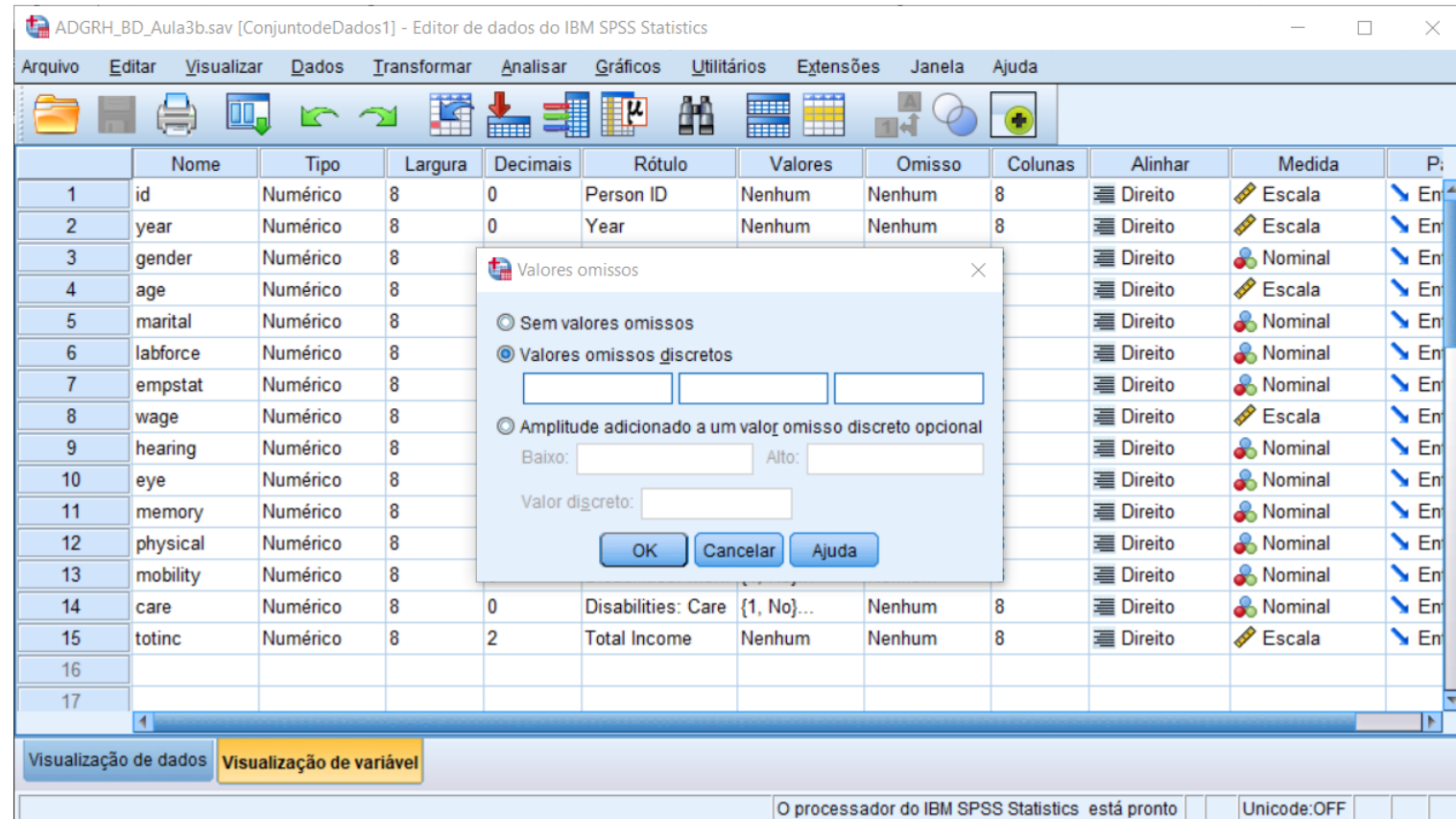
O processador do IBM SPSS Statistics está pronto Unicode:OFF

Definir Valores Omissos

- O SPSS oferece-nos uma maneira muito simples de fazer isso...
- Selecionar 'Visualização de variável'
- Na coluna 'Omisso', clicar na célula da variável 'labforce'

A

B



Definir Valores Omissos

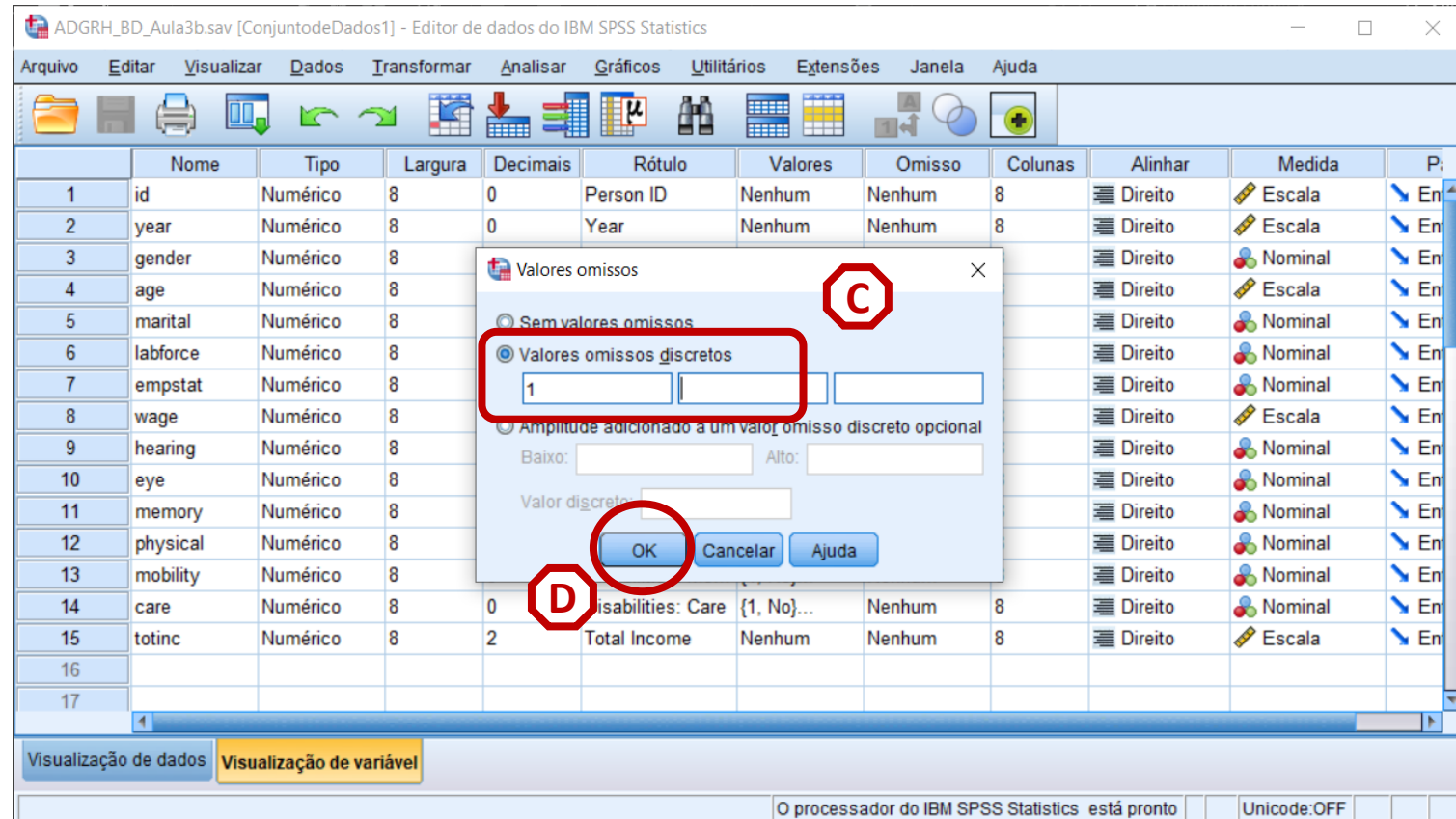
- O SPSS oferece-nos uma maneira muito simples de fazer isso...
- Selecionar 'Visualização de variável'
- Na coluna 'Omisso', clicar na célula da variável 'labforce'
- Selecionar 'Valores omissos discretos', e instruir o SPSS que o valor 1 deve ser tratado como valor omissso.
- Clicar 'OK'

A

B

C

D



Definir Valores Omissos

- O SPSS oferece-nos uma maneira muito simples de fazer isso...
- Selecionar 'Visualização de variável'
- Na coluna 'Omisso', clicar na célula da variável 'labforce'
- Selecionar 'Valores omissos discretos', e instruir o SPSS que o valor 1 deve ser tratado como valor omisso.
- Clicar 'OK'
- Clicar o ícone 'Salvar este documento'



A

B

C

D

E

*ADGRH_BD_Aula3b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF

		Tipo	Largura	Decimais	Rótulo	Valores	Omisso	Colunas	Alinhar	Medida	P:
1	id	Númerico	8	0	Person ID	Nenhum	Nenhum	8	Direito	Escala	En
2	year	Númerico	8	0	Year	Nenhum	Nenhum	8	Direito	Escala	En
3	gender	Númerico	8	0	Gender	{1, Male}...	Nenhum	8	Direito	Nominal	En
4	age	Númerico	8	2	Age	Nenhum	Nenhum	8	Direito	Escala	En
5	marital	Númerico	8	0	Marital Status	{1, Married, ...	Nenhum	8	Direito	Nominal	En
6	labforce	Númerico	8	0	In the Labour F...	{1, Not in U...	1	8	Direito	Nominal	En
7	empstat	Númerico	8	0	Employment St...	{1, Employe...	Nenhum	8	Direito	Nominal	En
8	wage	Númerico	8	2	Wage	Nenhum	Nenhum	8	Direito	Escala	En
9	hearing	Númerico	8	0	Disabilities: He...	{1, No}...	Nenhum	8	Direito	Nominal	En
10	eye	Númerico	8	0	Disabilities: Ey...	{1, No}...	Nenhum	8	Direito	Nominal	En
11	memory	Númerico	8	0	Disabilities: Me...	{1, No}...	Nenhum	8	Direito	Nominal	En
12	physical	Númerico	8	0	Disabilities: Ph...	{1, No}...	Nenhum	8	Direito	Nominal	En
13	mobility	Númerico	8	0	Disabilities: Mo...	{1, No}...	Nenhum	8	Direito	Nominal	En
14	care	Númerico	8	0	Disabilities: Care	{1, No}...	Nenhum	8	Direito	Nominal	En
15	totinc	Númerico	8	2	Total Income	Nenhum	Nenhum	8	Direito	Escala	En
16											
17											

Limpar a Base de Dados

Recodificar ‘Valores Omissos’

Objetivo: Criar uma nova variável ‘wage2’ em que o valor ‘999999999’ é codificado como system missing.

**8 noves*

Recodificar como Valor Omisso

- Na variável 'wage' os valores omissos estão codificados como:

99999999



- Objectivo

- Criar uma nova variável 'wage2' em que o valor '99999999' é codificado como system missing.

*ADGRH_BD_Aula3b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Visível: 15 de 15 variáveis

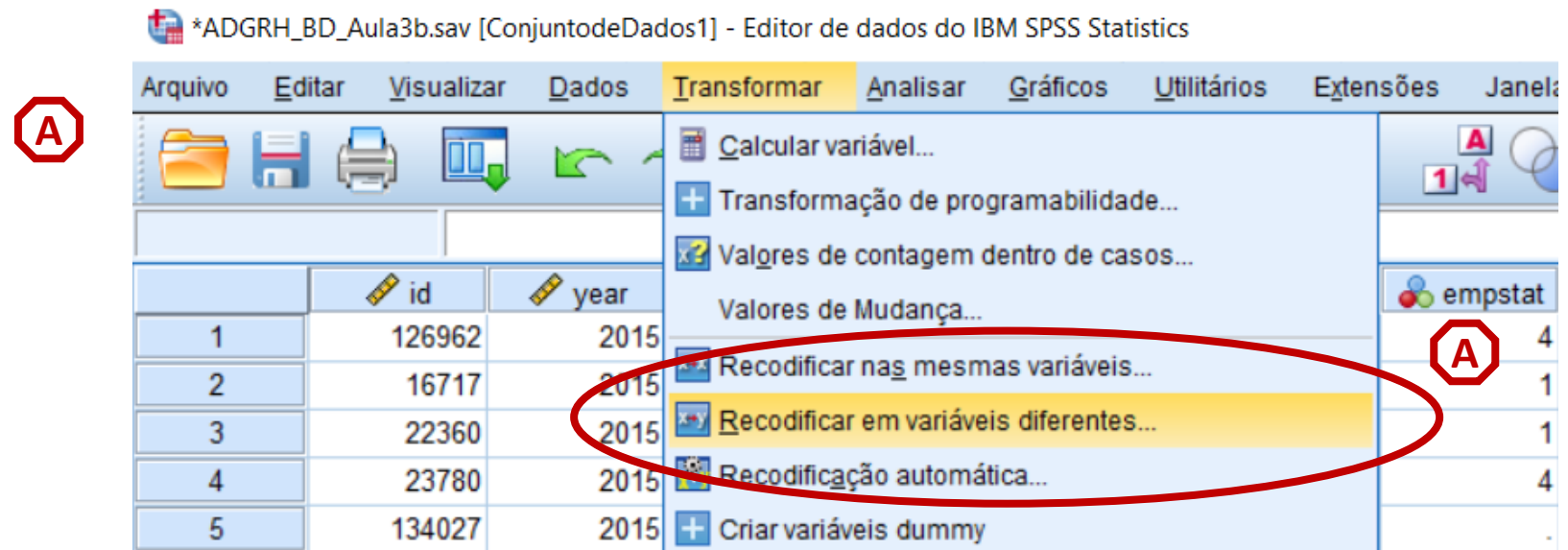
	id	year	gender	age	marital	labforce	empstat	wage	hearing	eye	memory
1	126962	2015	1	52,00	4	2	4	99999999,00	1	1	1
2	16717	2015	1	32,00	1	3	1	6141,00	1	1	1
3	22360	2015	2	46,00	1	3	1	46000,00	1	1	1
4	23780	2015	1	44,00	1	1	4	120000,00	.	.	.
5	134027	2015	1	12,00	6	1	4	99999999,00	.	.	.
6	13215	2015	1	21,00	6	2	4	6500,00	1	1	1
7	173046	2015	2	52,00	6	2	3	99999999,00	1	1	1
8	195945	2015	1	8,00	6	1	.	99999999,00	.	.	.
9	97355	2015	2	18,00	6	3	1	7002,00	1	1	1
10	180306	2015	2	15,00	6	3	2	99999999,00	1	1	1
11	170018	2015	1	67,00	4	3	1	99999999,00	2	1	1
12	159824	2015	1	66,00	5	2	5	99999999,00	1	1	1
13	67213	2015	2	,00	6	1	.	99999999,00	.	.	.
14	100988	2015	1	29,00	4	3	1	20700,00	1	1	1
15	18096	2015	2	14,00	6	1	.	99999999,00	.	.	.
16	107715	2015	2	40,00	4	2	1	10000,00	1	1	1

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF

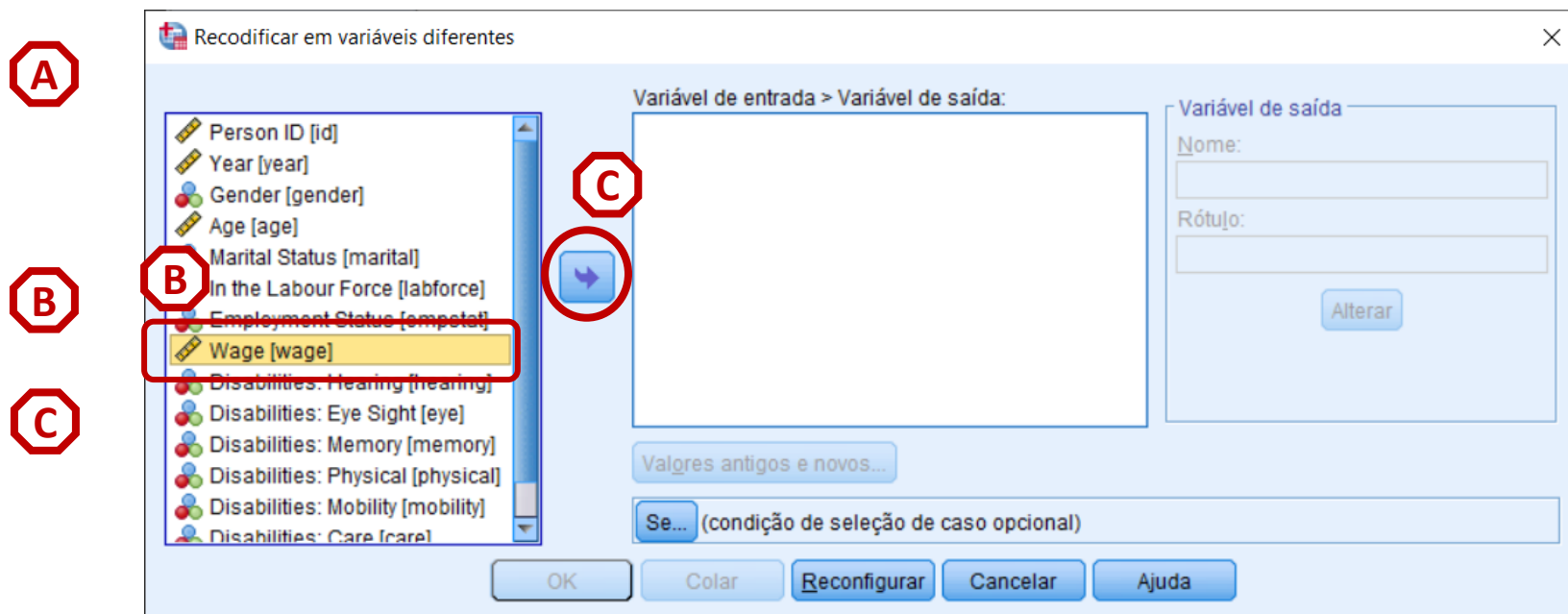
Recodificar como Valor Omisso

- Selecionar 'Transformar' / 'Recodificar em variáveis diferentes'



Recodificar como Valor Omisso

- Selecionar 'Transformar' / 'Recodificar em variáveis diferentes'
- Selecionar a variável 'wage'
- Colocá-la na caixa 'Variável numérica'



Recodificar como Valor Omisso

- Selecionar 'Transformar' / 'Recodificar em variáveis diferentes'
- Selecionar a variável 'wage'
- Colocá-la na caixa 'Variável numérica'
- Definir o nome da variável como 'wage2' e o rótulo como 'Wage'
- Clicar 'Alterar'
- Clicar e 'Valores antigos e novos'

A

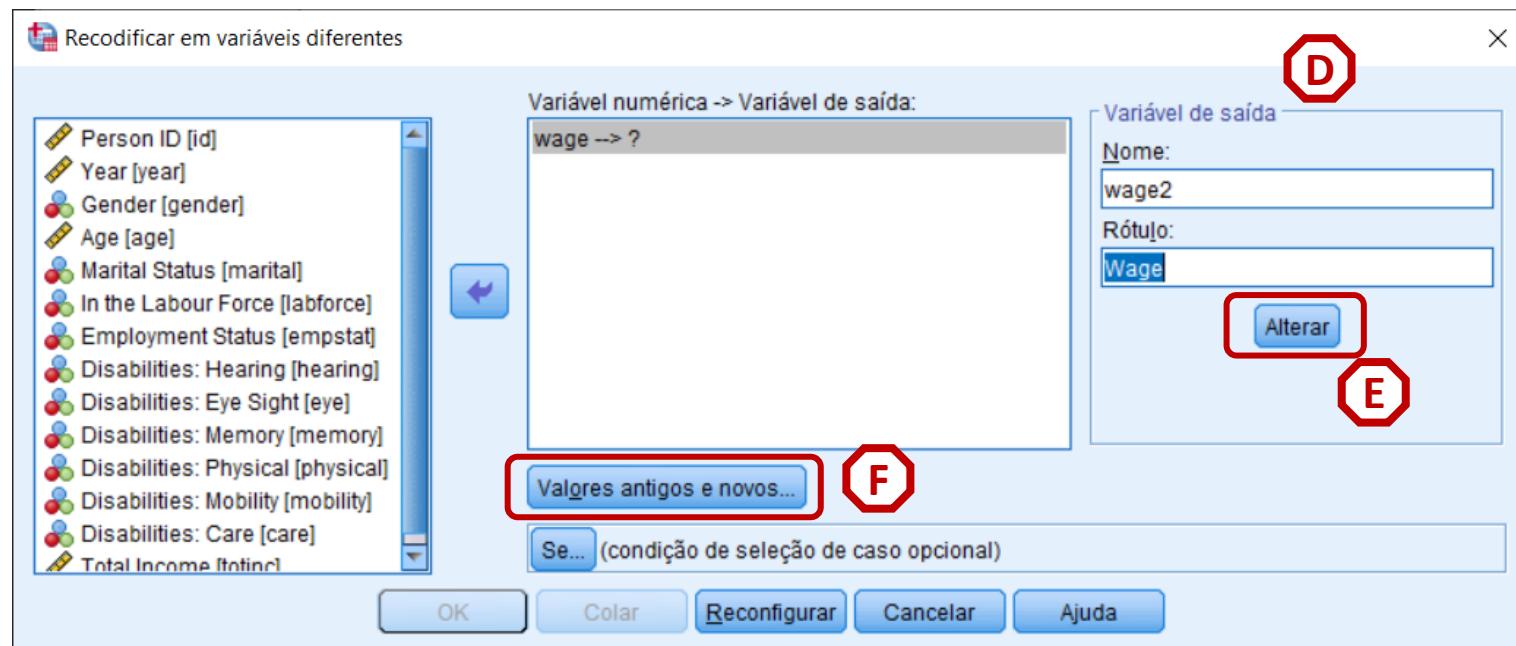
B

C

D

E

F



Recodificar como Valor Omisso

- Primeiro, instruímos o SPSS que todas as observações com o valor '99999999'
... devem ser tratadas como 'System- missing' na nova variável
- Clicar 'Incluir'

G

H

I

Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo

☒ Valor: 99999999

☐ Omisso no sistema

☐ Omisso para o sistema ou usuário

☐ Amplitude: até

☐ Intervalo, MAIS BAIXO até o valor:

☐ Intervalo, valor até o MAIS ALTO:

☒ Todos os outros valores

Novo valor

☐ Valor:

☒ Omisso no sistema

☐ Copiar valor(es) antigo(s)

Antigo --> Novo:

☐ As variáveis de saída são sequências de caracteres Largura: 8

☒ Converter sequências de caracteres numéricas em números ('5' -> 5)

Recodificar como Valor Omisso

- Primeiro, instruímos o SPSS que todas as observações com o valor '99999999'
- ... devem ser tratadas como 'System- missing' na nova variável
- Clicar 'Incluir'



Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo

☒ Valor:

☐ Omisso no sistema

☐ Omisso para o sistema ou usuário

☐ Amplitude:

até

☐ Intervalo, MAIS BAIXO até o valor:

☐ Intervalo, valor até o MAIS ALTO:

☒ Todos os outros valores

Novo valor

☐ Valor:

☒ Omisso no sistema

☐ Copiar valor(es) antigo(s)

Antigo --> Novo:

99999999 --> SYSMIS

Incluir

Alterar

Remover

☐ As variáveis de saída são sequências de caracteres Largura:

☒ Converter sequências de caracteres numéricas em números ('5' -> 5)

Continuar Cancelar Ajuda

Recodificar como Valor Omisso

- Primeiro, instruímos o SPSS que todas as observações com o valor '99999999'

... devem ser tratadas como 'System- missing' na nova variável

- Clicar 'Incluir'

- Depois, definir que 'Todos os outros valores'

... deve ser copiados ('Copiar valor(es) antigo(s)') para a nova variável

- Clicar 'Incluir'

Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo

- ☐ Valor:
- ☐ Omisso no sistema
- ☐ Omisso para o sistema ou usuário
- ☐ Amplitude:
- ☐ Intervalo, MAIS BAIXO até o valor:
- ☐ Intervalo, valor até o MAIS ALTO:
- ☒ Todos os outros valores

Novo valor

- ☐ Valor:
- ☐ Omisso no sistema
- ☒ Copiar valor(es) antigo(s)

Antigo --> Novo:
99999999 --> SYSMIS

Incluir
Alterar
Remover

As variáveis de saída são sequências de caracteres Largura: 8
Converter sequências de caracteres numéricas em números ('5'-->5)

Continuar Cancelar Ajuda

Recodificar como Valor Omisso

- Primeiro, instruímos o SPSS que todas as observações com o valor '99999999'
... devem ser tratadas como 'System-missing' na nova variável
- Clicar 'Incluir'
- Depois, definir que 'Todos os outros valores'
... deve ser copiados ('Copiar valor(es) antigo(s)') para a nova variável
- Clicar 'Incluir'
- Clicar 'Continuar' / 'OK'

G

H

I

J

K

L

M

Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo

- ☐ Valor:
- ☐ Omisso no sistema
- ☐ Omisso para o sistema ou usuário
- ☐ Amplitude:
- ☐ Intervalo, MAIS BAIXO até o valor:
- ☐ Intervalo, valor até o MAIS ALTO:
- ☒ Todos os outros valores

Novo valor

- ☐ Valor:
- ☐ Omisso no sistema
- ☒ Copiar valor(es) antigo(s)

Antigo --> Novo:
99999999 --> SYSMIS
ELSE --> Copy

Incluir
Alterar
Remover

☐ variáveis de saída são sequências de caracteres Largura: 8
☐ Inverter sequências de caracteres numéricas em números ('5'-->5)

Continuar Cancelar Ajuda

Recodificar como Valor Omisso

- Vamos verificar

*ADGRH_BD_Aula3b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : wage 99999999,00 Visível: 16 de 16 variáveis

	mpstat	wage	hearing	eye	memory	physical	mobility	care	totinc	wage2	var
1	4	99999999,00	1	1	1	1	1	1	.	.	
2	1	6141,00	1	1	1	1	1	1	39156,00	6141,00	
3	1	46000,00	1	1	1	1	1	1	46277,00	46000,00	
4	4	120000,00	.	.	.	.	.	.	130300,00	120000,00	
5	.	99999999,00	.	.	.	.	.	.	.	.	
6	4	6500,00	1	1	1	1	1	1	.	6500,00	
7	3	99999999,00	1	1	1	1	1	1	.	.	
8	.	99999999,00	.	.	.	.	.	.	.	.	
9	1	7002,00	1	1	1	1	1	1	.	7002,00	
10	2	99999999,00	1	1	1	1	1	1	.	.	
11	1	99999999,00	2	1	1	1	1	1	51812,00	.	
12	5	99999999,00	1	1	1	1	1	1	36663,00	.	
13	.	99999999,00	.	.	.	.	.	.	.	.	
14	1	20700,00	1	1	1	1	1	1	.	20700,00	
15	.	99999999,00	.	.	.	.	.	.	.	.	
16	1	10000,00	1	1	1	1	1	1	.	10000,00	

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF

Limpar a Base de Dados

Substituir o Valor Omisso por Outro Valor

Objetivo: Recodificar a variável ‘wage2’ numa nova variável em que os valores omissos são substituídos pela média da variável .

Substituir o Valor Omisso pela Média

- A variável 'wage2' tem valores omissos (system-missing).
- Objectivo
 - Recodificar a variável 'wage2' numa nova variável em que os valores omissos são substituídos pela média da variável*

* Também podemos substituir pela mediana, por exemplo...



*ADGRH_BD_Aula3b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

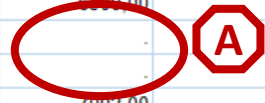
Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

29 : Visível: 16 de 16 variáveis

	mpstat	wage	hearing	eye	memory	physical	mobility	care	totinc	wage2	var
1	4	9999999,00	1	1	1	1	1	1	.	.	
2	1	6141,00	1	1	1	1	1	1	39156,00	6141,00	
3	1	46000,00	1	1	1	1	1	1	46277,00	46000,00	
4	4	120000,00	.	.	.	.	.	.	130300,00	120000,00	
5	.	99999999,00	.	.	.	.	.	.	.	.	
6	4	6500,00	1	1	1	1	1	1	.	6500,00	
7	3	99999999,00	1	1	1	1	1	1	.	.	
8	.	99999999,00	.	.	.	.	.	.	.	.	
9	1	7002,00	1	1	1	1	1	1	.	7002,00	
10	2	99999999,00	1	1	1	1	1	1	.	.	
11	1	99999999,00	2	1	1	1	1	1	51812,00	.	
12	5	99999999,00	1	1	1	1	1	1	36663,00	.	
13	.	99999999,00	.	.	.	.	.	.	.	.	
14	1	20700,00	1	1	1	1	1	1	.	20700,00	
15	.	99999999,00	.	.	.	.	.	.	.	.	
16	1	10000,00	1	1	1	1	1	1	.	10000,00	

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF



Substituir o Valor Omisso pela Média

- Selecionar 'Transformar' /
'Substituir valores omissos'



*ADGRH_BD_Aula3b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

29 :

	mpstat	wage	h	mobility	care
1	4	99999999,00		1	1
2	1	6141,00		1	1
3	1	46000,00		1	1
4	4	120000,00		.	.
5	.	99999999,00		.	.
6	4	6500,00		1	1
7	3	99999999,00		1	1
8	.	99999999,00		.	.
9	1	7002,00		1	1
10	2	99999999,00		1	1
11	1	99999999,00		1	1
12	5	99999999,00		1	1
13	.	99999999,00		.	.
14	1	20700,00		1	1

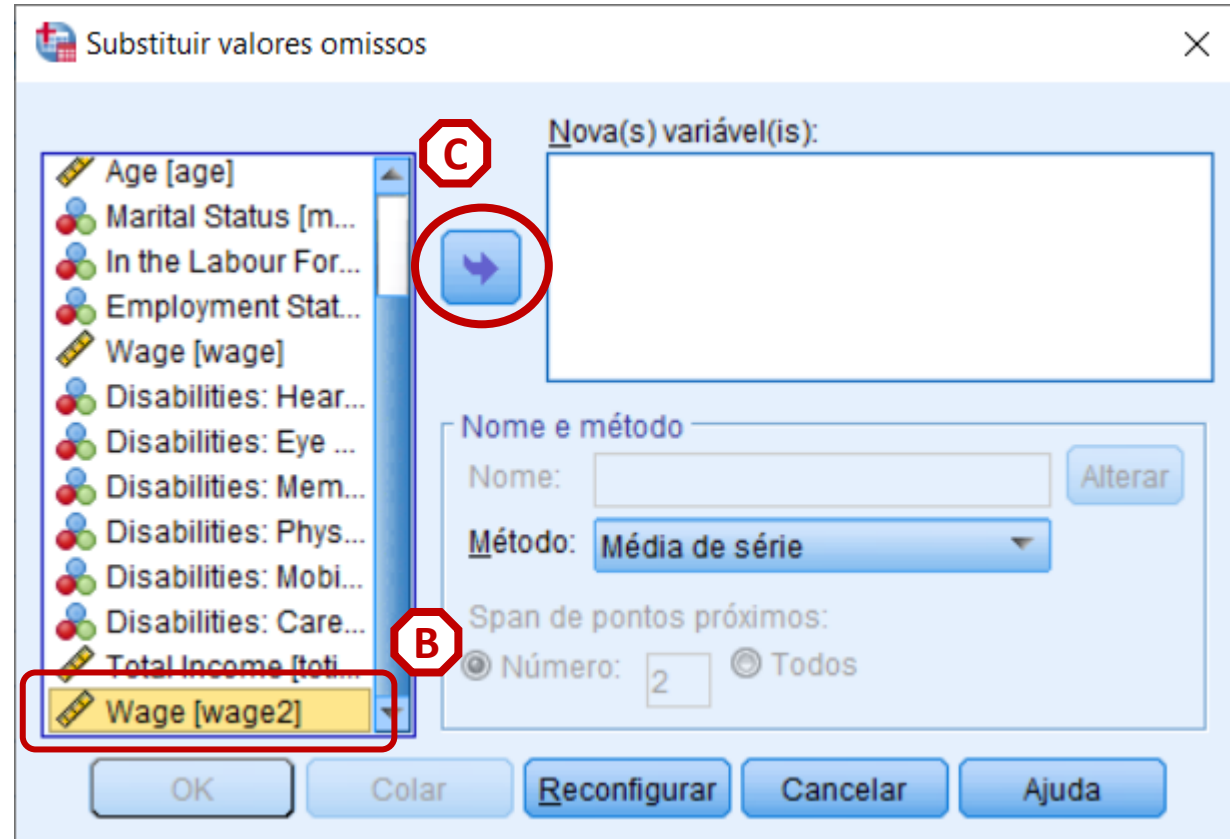
Transformar

- Calcular variável...
- Transformação de programabilidade...
- Valores de contagem dentro de casos...
- Valores de Mudança...
- Recodificar nas mesmas variáveis...
- Recodificar em variáveis diferentes...
- Recodificação automática...
- Criar variáveis dummy
- Categorização visual...
- Categorização ideal...
- Preparar dados para modelagem
- Planquear Casos...
- Assistente de data e hora...
- Criar séries temporais...
- Substituir valores omissos...**
- Geradores de número aleatório...

Red octagonal icon with the letter A

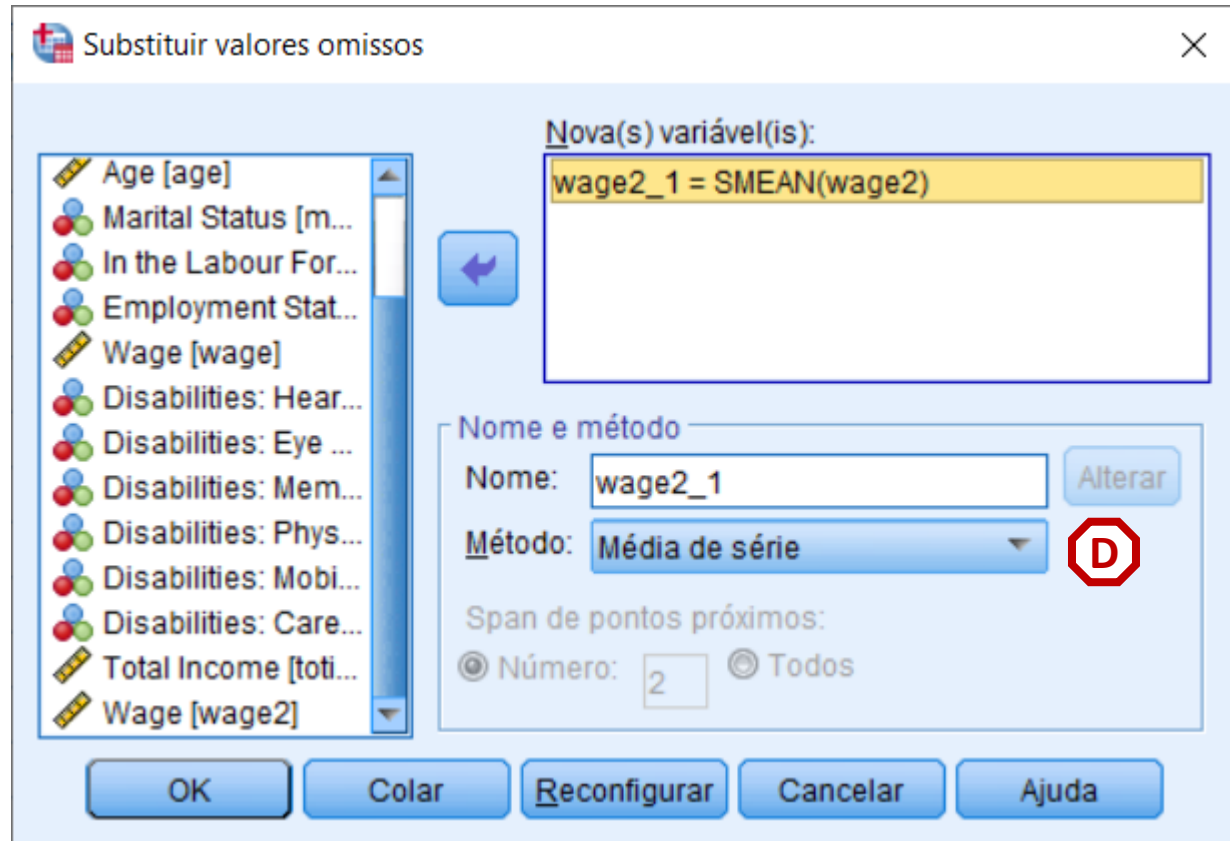
Substituir o Valor Omisso pela Média

- Selecionar 'Transformar' /
'Substituir valores omissos' **A**
- Selecionar a variável 'wage2' **B**
- Colocá-la na caixa 'Nova(s) Variável(is)' **C**



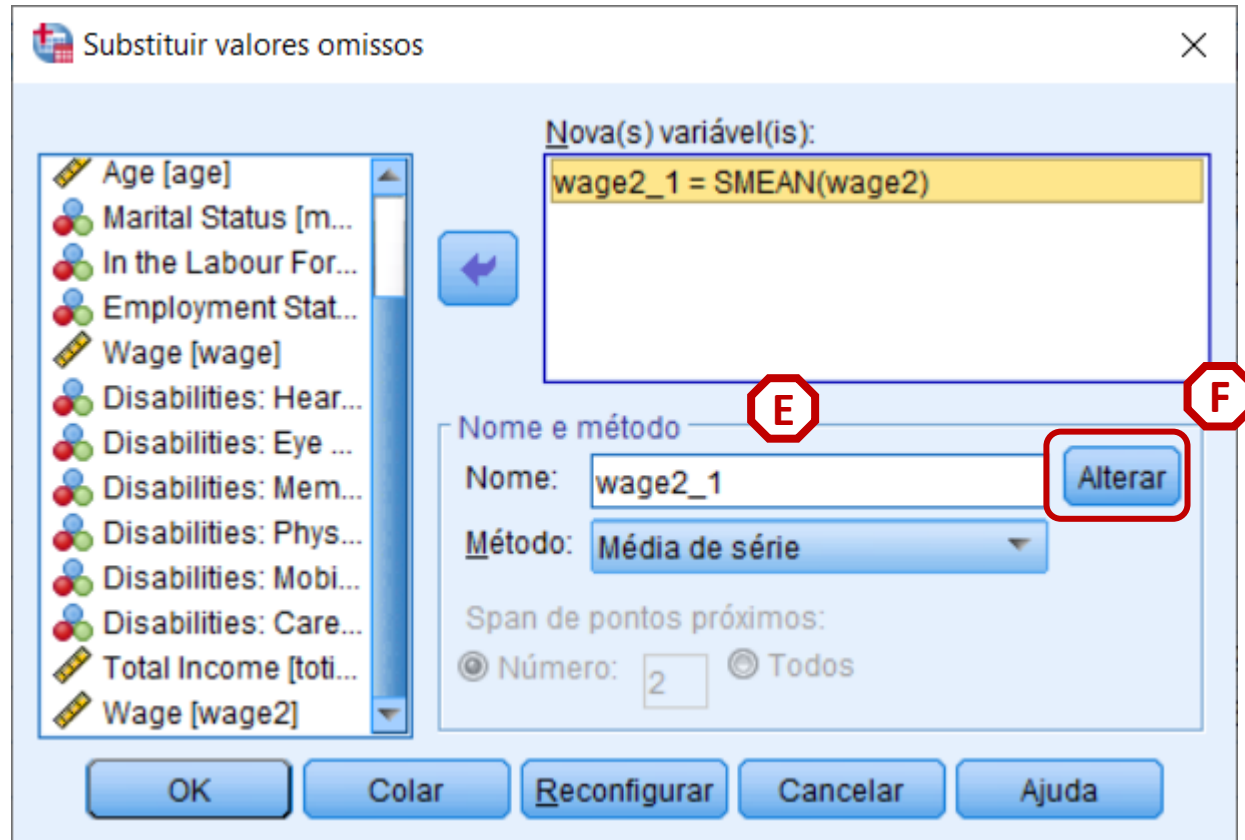
Substituir o Valor Omisso pela Média

- Selecionar 'Transformar' / 'Substituir valores omissos' **A**
- Selecionar a variável 'wage2' **B**
- Colocá-la na caixa 'Nova(s) Variável(is)' **C**
- Por defeito, o 'Método' de substituição é 'Média de Série' **D**



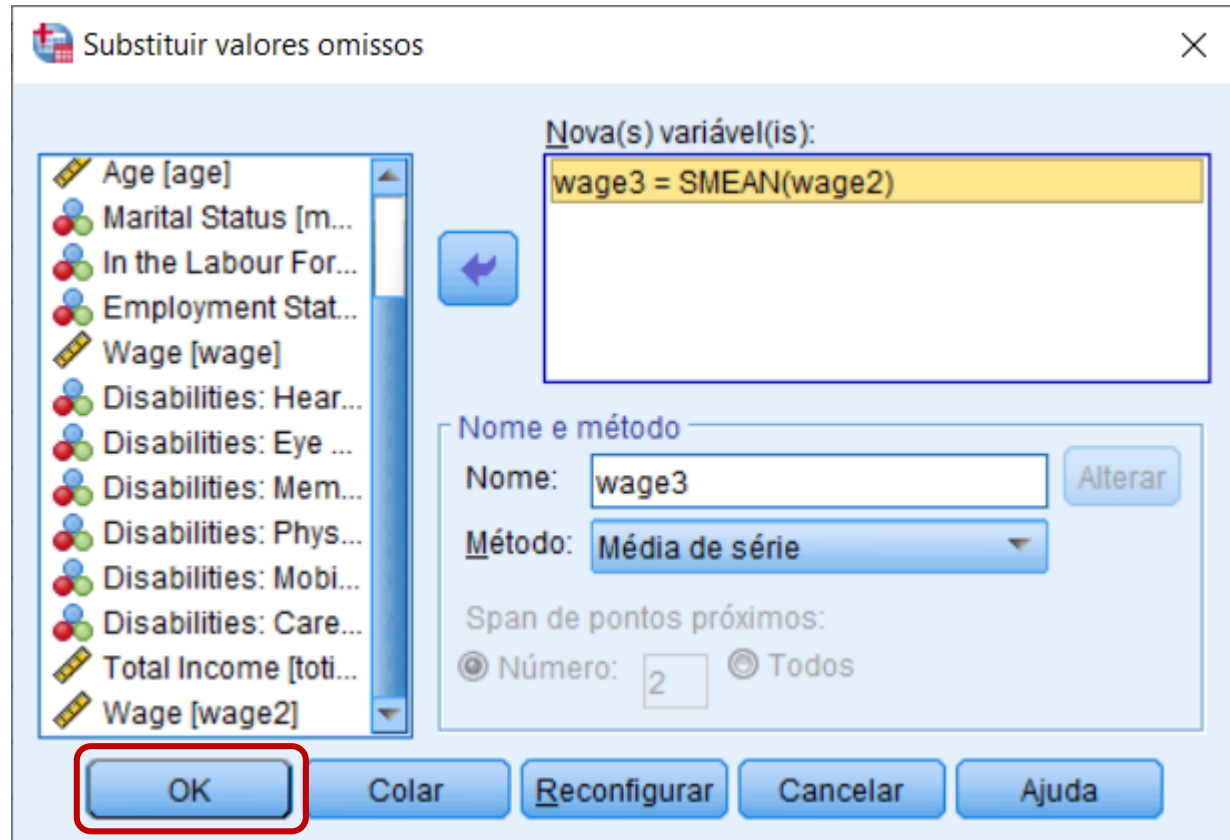
Substituir o Valor Omisso pela Média

- Selecionar 'Transformar' / 'Substituir valores omissos' **A**
- Selecionar a variável 'wage2' **B**
- Colocá-la na caixa 'Nova(s) Variável(is)' **C**
- Por defeito, o 'Método' de substituição é 'Média de Série' **D**
- Mudar o nome para 'wage3' **E**
- Clicar 'Alterar' **F**



Substituir o Valor Omisso pela Média

- Selecionar 'Transformar' / 'Replace Missing Values' **A**
- Selecionar the 'wage2' variable **B**
- Colocá-la na caixa 'Nova(s) Variável(is)' **C**
- Por defeito, o 'Método' de substituição é 'Média de Série' **D**
- Mudar o nome para 'wage3' **E**
- Clicar 'Alterar' **F**
- Clicar 'OK' **G**



Substituir o Valor Omisso pela Média

- Podemos ver o número de valores omissos foram recodificados...
- Na nova variável todas as 5000 observações têm dados sobre o salário

A

B

*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Inserir Formatar Analisar Gráficos Utilitários Extensões Janela Ajuda

Estatística Descritiva

	N	Mínimo	Máximo	Média	Erro Desvio
Wage	2194	10,00	250000,00	42572,8888	36984,22846
N válido (de lista)	2194				

→ Substituir valores omissos

RECODE wage (99999999=SYSMIS) (ELSE=Copy) INTO wage2.
VARIABLE LABELS wage2 'Wage'.
EXECUTE.
RMV /wage3=SMEAN(wage2).

Variáveis de Resultado

	Variável de Resultado	N de Valores Omissos Substituídos	Número de Caso de Valores Não omissos		N de Casos Válidos	Criando Função
			Primeiro	Último		
1	wage3	2806	1	5000	5000	SMEAN (wage2)

DESCRIPTIVES VARIABLES=wage3
/STATISTICS=MEAN STDDEV MIN MAX.

→ Descritivos

Estatística Descritiva

	N	Mínimo	Máximo	Média	Erro Desvio
SMEAN(wage2)	5000	10,00	250000,00	42572,8888	24495,95134
N válido (de lista)	5000				

Visualização prévia da impressão deste conjunto de dados

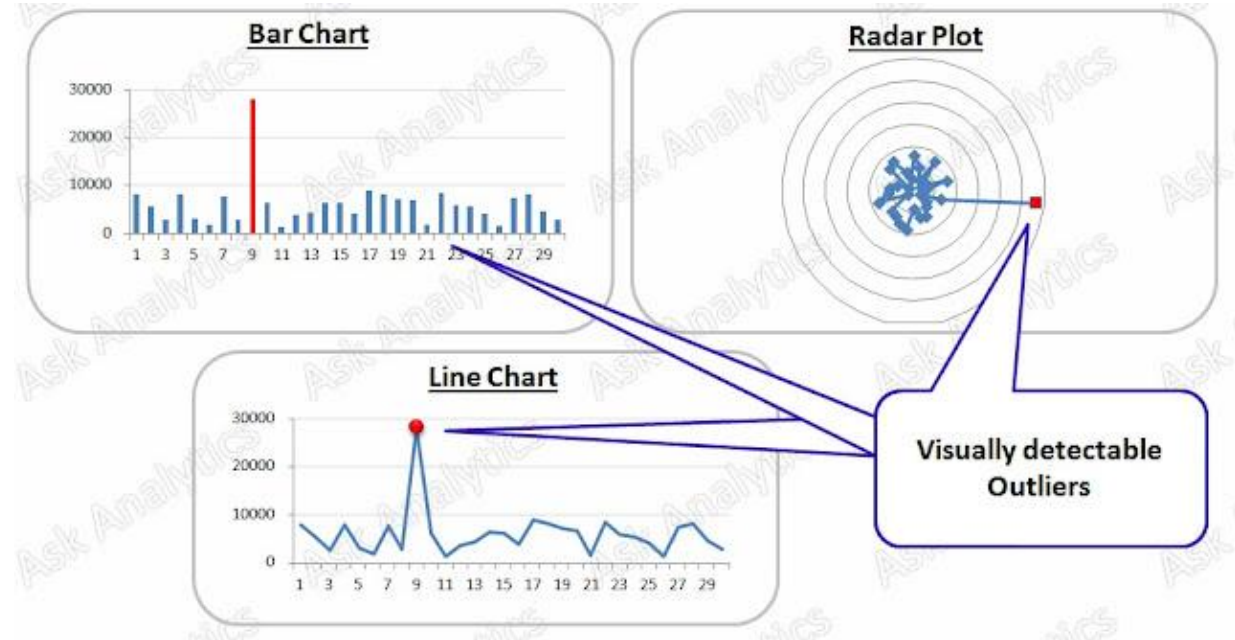
O processador do IBM SPSS Statistics está pronto | Unicode:OFF | H: 2,12, W: 33,52 cm

67

O que são ‘valores extremos’?

=“outliers”

- São observações que se desviam de uma forma muito significativa do padrão de distribuição das observações de uma variável.
- Só acontecem em variáveis intervalares e de razão.



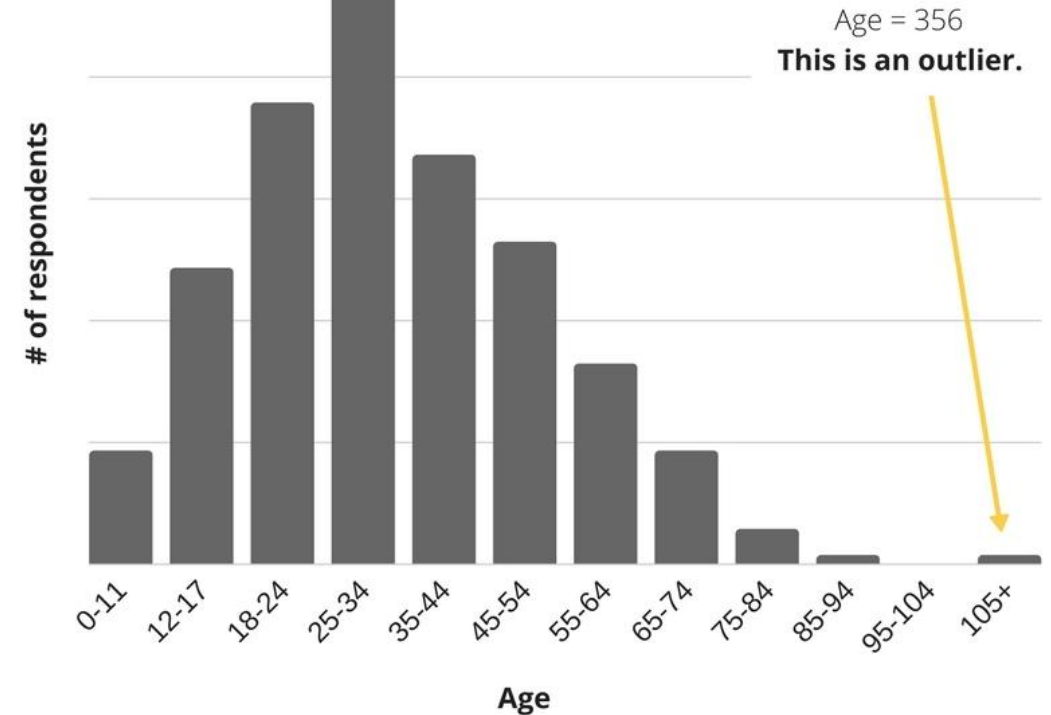
Fonte: <http://www.askanalytics.in/p/outl.html>

- **O que é que pode explicar a existência de valores extremos?**
 - **Valores extremos naturais (ex: Salário do CEO)**
 - **Resultantes de erros na recolha e tratamento de dados**
 - Respostas incorrectas (ex: mencionar um salário muito alto/baixo para seconder o verdadeiro valor).
 - Erros de medição (ex: avaria no dispositivo de recolha de dados biométricos).
 - Erros na imputação de dados (ex: adicionar um 0 a mais) *

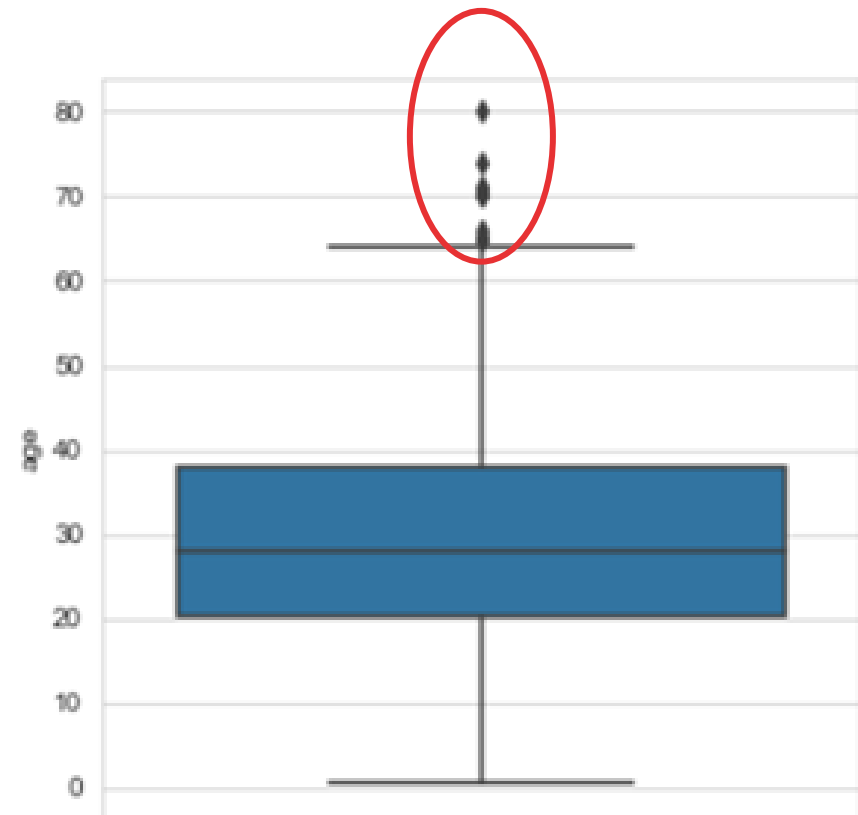
*** Nos casos em que valores omissos codificados como '99999' (ou outras sequências), estes não devem ser tratados como valores extremos.**

- **Por que é que a existência de 'valores extremos' é um problema?**
 - **Enviesam medidas básicas de estatística descritiva (médias, variância, desvio-padrão)**
 - **Têm implicações profundas nas medidas de inferência estatística:**
 - **Aumentam a variância e a margem de erro;**
 - **Podem enviesar os resultados dos testes estatísticos;**
 - **Podem significar a violação dos pressupostos de modelos de regressão linear.**

- Como é que podemos determinar a possibilidade de ‘valores extremos’ numa variável?
- Visualização da distribuição das variáveis:
 - Histograma (menos útil)



- **Como é que podemos determinar a possibilidade de ‘valores extremos’ numa variável?**
- **Visualização da distribuição das variáveis:**
 - **Histograma (menos útil)**
 - **Gráfico de Caixa**



- **Como é que podemos determinar a possibilidade de ‘valores extremos’ numa variável?**
 - **Visualização da distribuição das variáveis:**
 - **Histograma**
 - **Gráfico de Caixa**
 - **Aplicando um critério estatístico (z-score)**

- **O que fazer se temos ‘valores extremos’ na nossa variável?**
 - **Em primeiro lugar, devemos determinar se se trata de um outlier natural, ou resulta de um erro.**
 - **Se determinarmos que estes valores extremos refletem de erros, temos duas alternativas:**
 - Recodificar os valores extremos como valores omissos (deve ser evitado, especialmente quando o tamanho da amostra é pequeno);
 - Limitar(/censurar) os valores nos extremos da distribuição (*‘winsorizing’*).

Limpar a Base de Dados

Identificar/Recodificar ‘Valores Extremos’

Diagnóstico Visual

Usar Z-Scores para Determinar ‘Valores Extremos’

Inspeccionar Potenciais ‘Valores Extremos’

Recodificar como Valor Omisso

Limpar a Base de Dados

Identificar visualmente ‘Valores Extremos’

Objetivo: Perceber se existem valores extremos na variável que mede o salário anual dos trabalhadores (Y_wage)

.

Gráfico de Caixa

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'Boxplot'
- Duplo-clique no Gráfico de Caixa (simples)

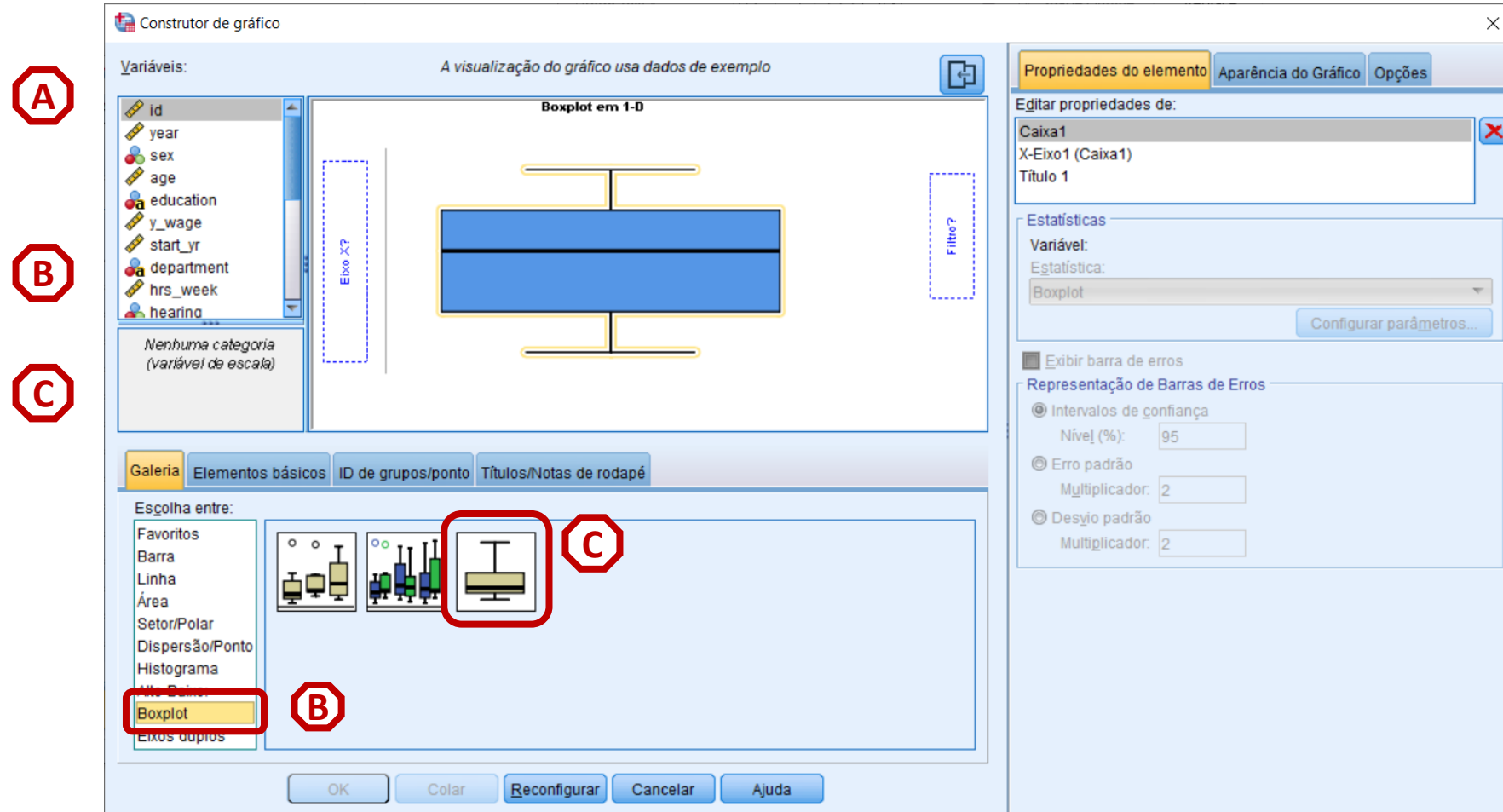


Gráfico de Caixa

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'Boxplot'
- Duplo-clique no Gráfico de Caixa (simples)
- Selecionar a variável 'Y_wage'
- E arrastar para o 'Eixo X'

A

B

C

D

E

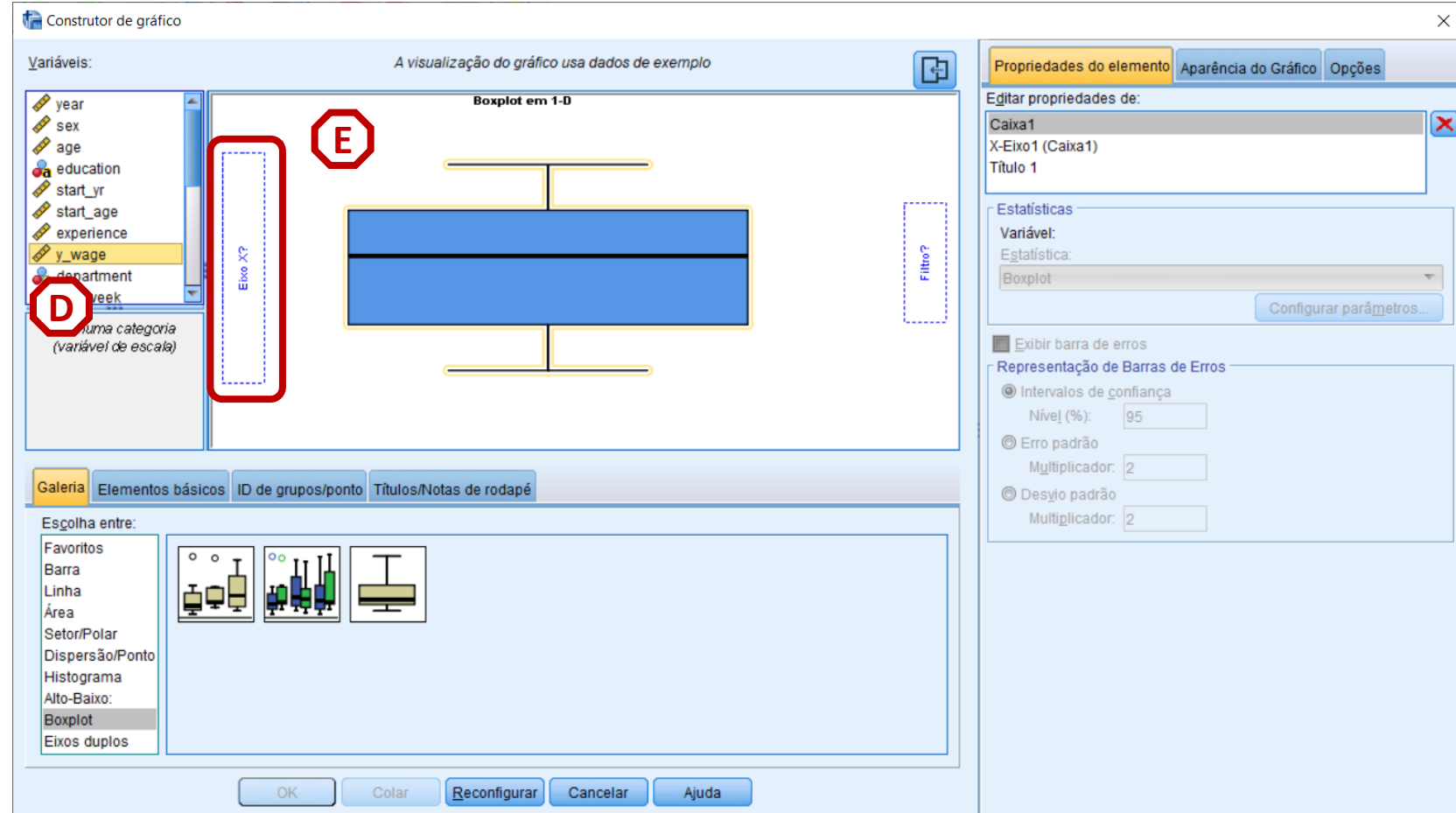


Gráfico de Caixa

- Selecionar 'Gráficos' / 'Construtor de Gráfico' **A**
- Selecionar 'Boxplot' **B**
- Duplo-clique no Gráfico de Caixa (simples) **C**
- Seleccionar a variável 'Y_wage' **D**
- E arrastar para o 'Eixo X' **E**
- Selecionar 'OK' **F**

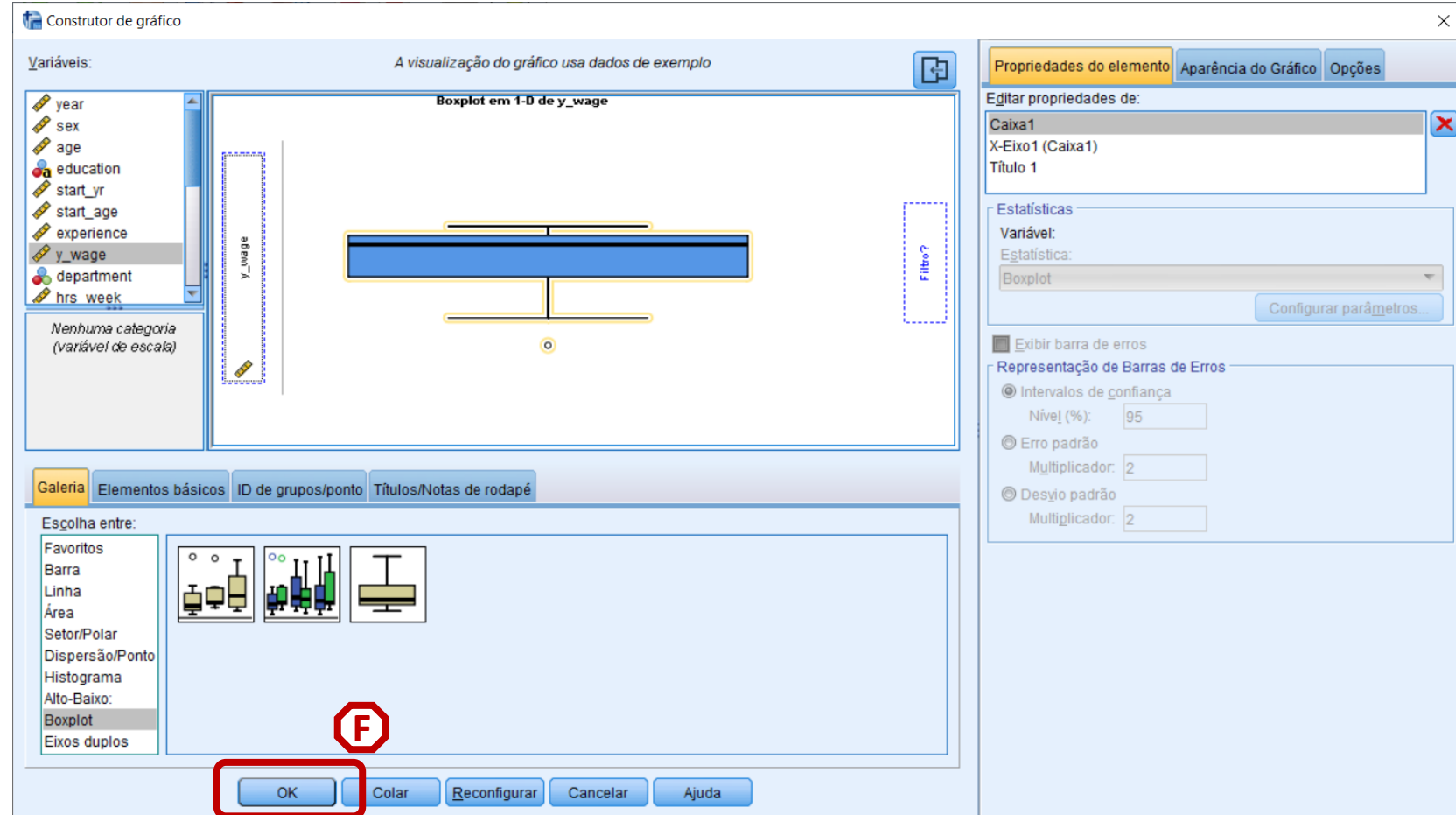


Gráfico de Caixa

- O gráfico é publicado no 'Visualizador de Resultados'

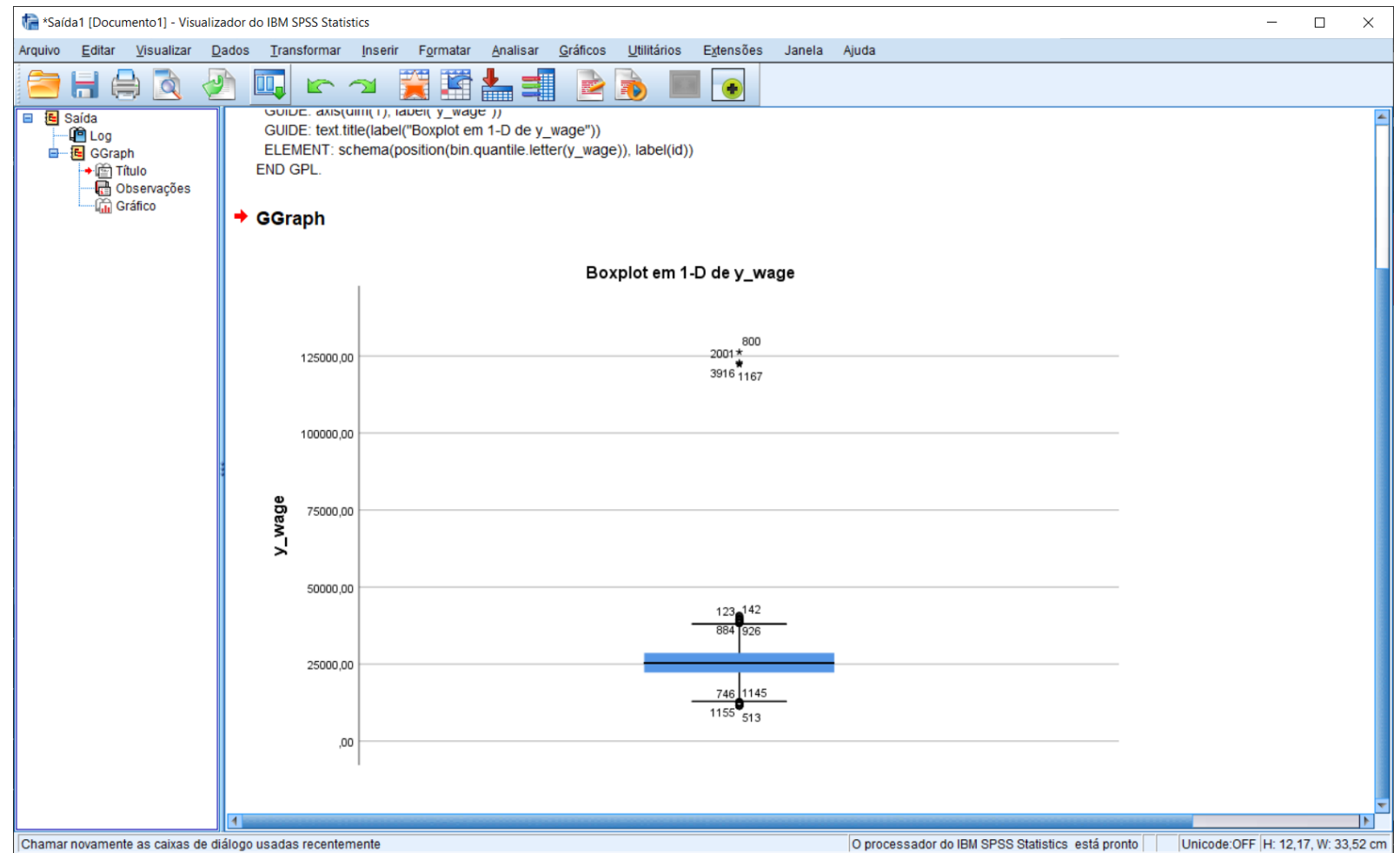
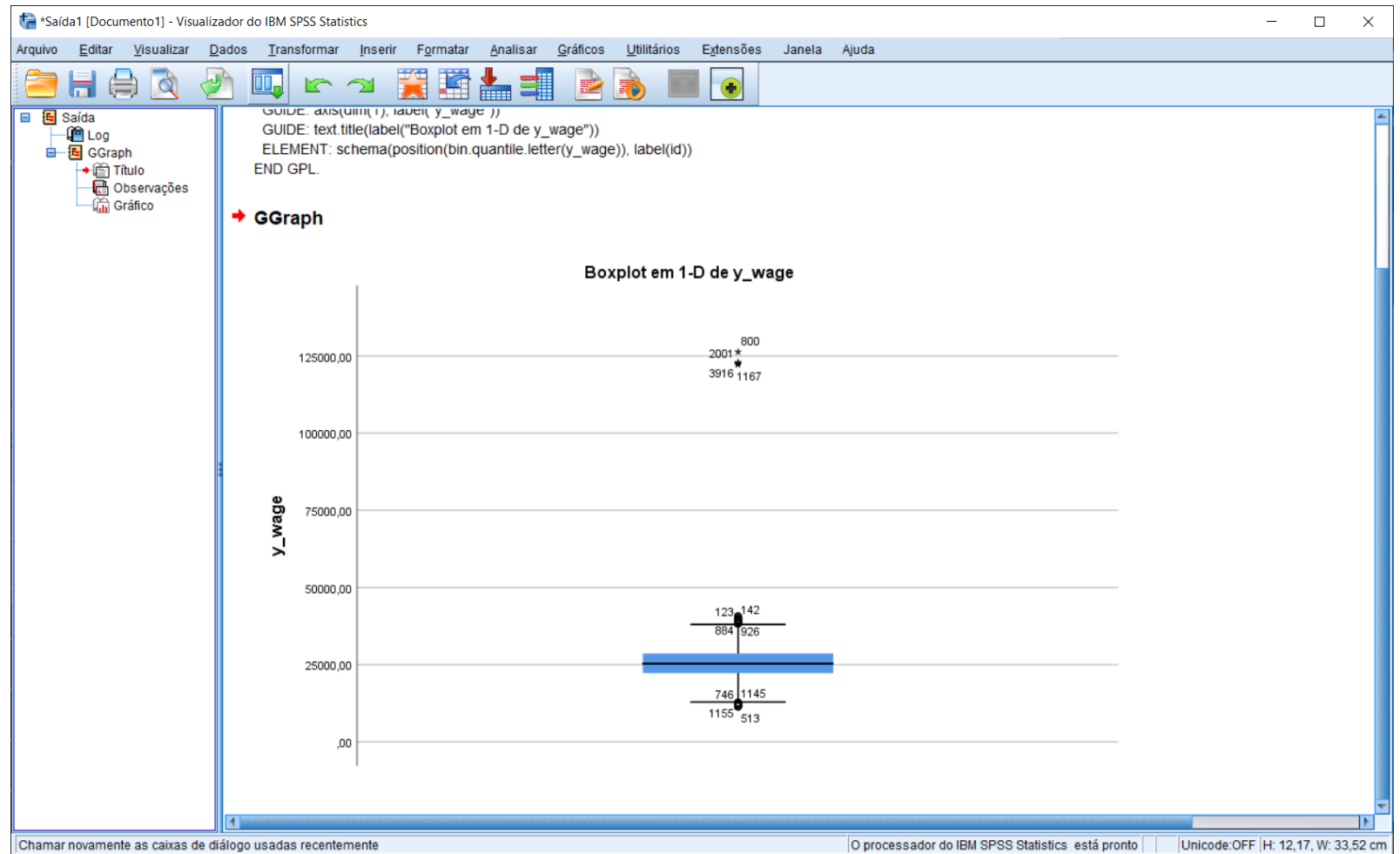


Gráfico de Caixa

- O gráfico é publicado no 'Visualizador de Resultados'

A representação gráfica da distribuição da variável sugere a existência de valores extremos.



Limpar a Base de Dados

Identificar/Recodificar 'Valores Extremos'

Usar Z-Scores para Determinar 'Valores Extremos'

Limpar a Base de Dados

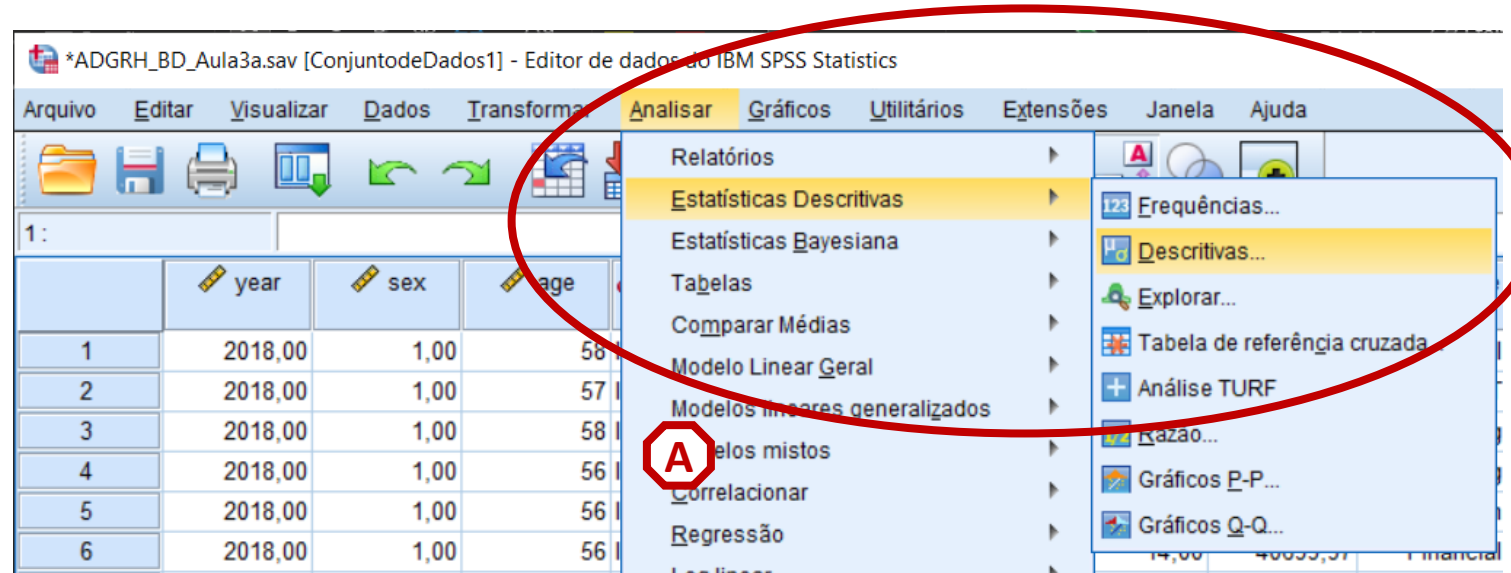
Identificar ‘Valores Extremos’

Objetivo: Usar Z-Scores para determinar ‘Valores Extremos’ na variável que mede o salário anual dos trabalhadores (Y_wage)

.

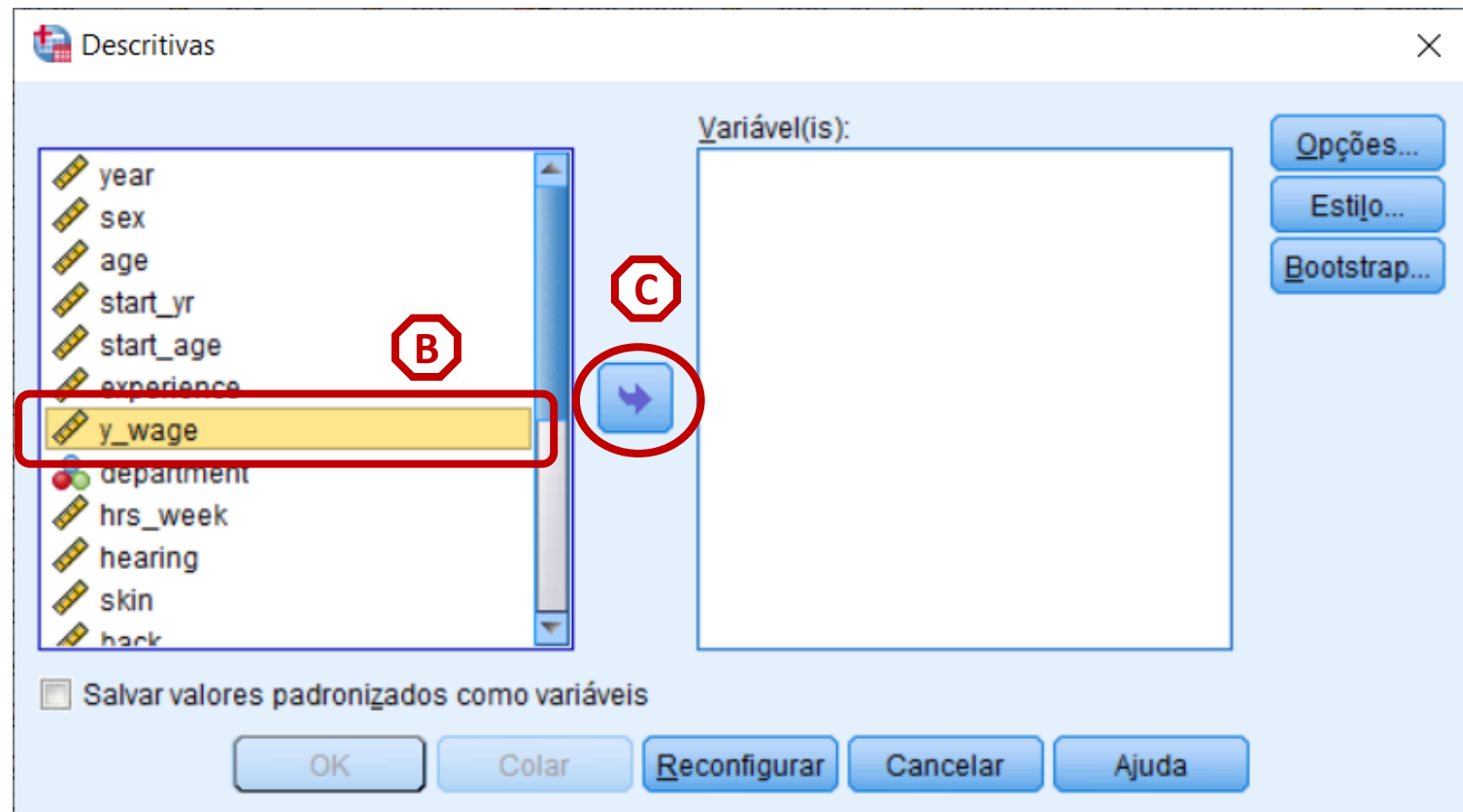
Z-Scores para identificar outliers

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Descritivas'



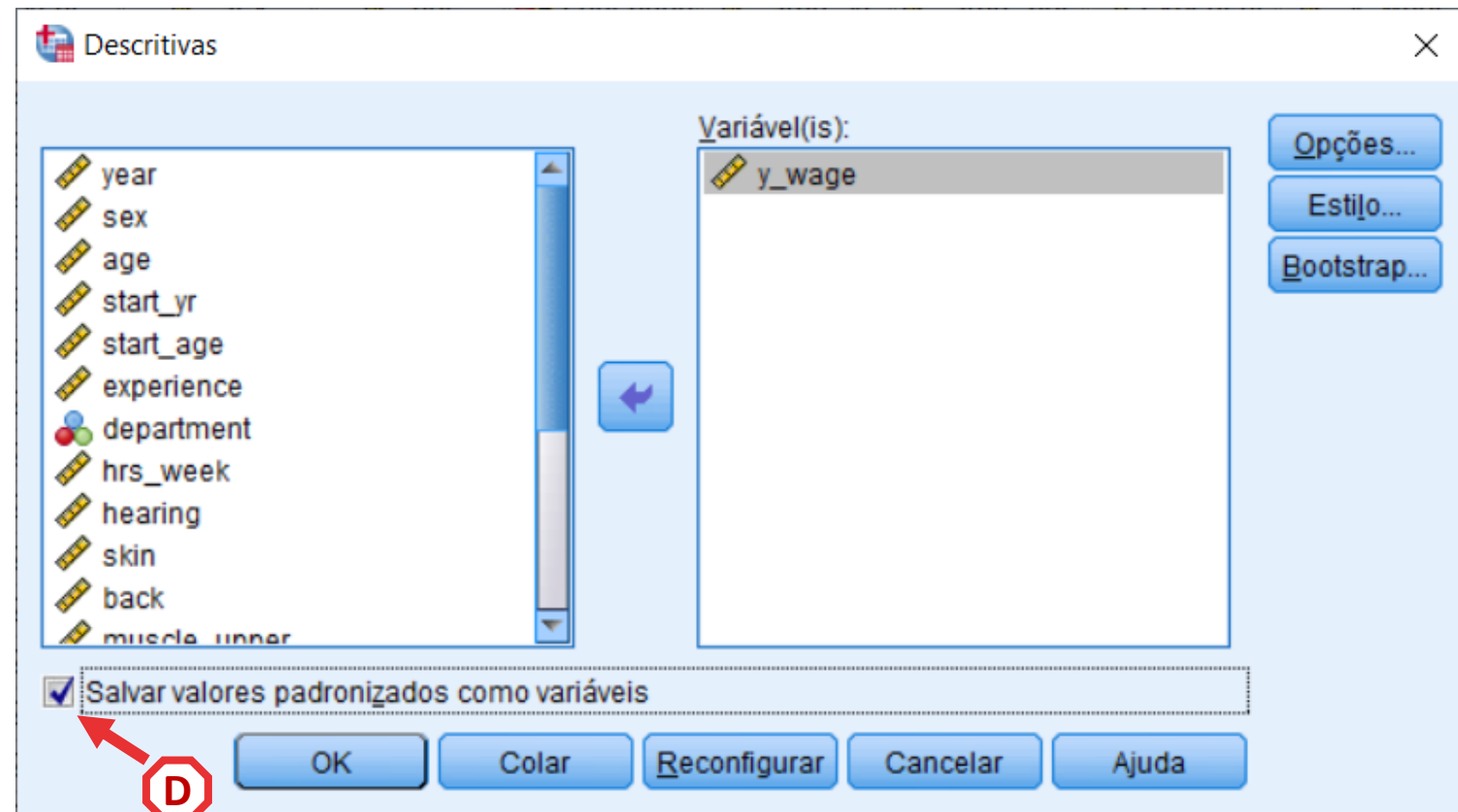
Z-Scores para identificar outliers

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Descritivas' **A**
- Selecionar a variável 'y_wage' **B**
- Colocar na caixa 'Variável(is)' **C**



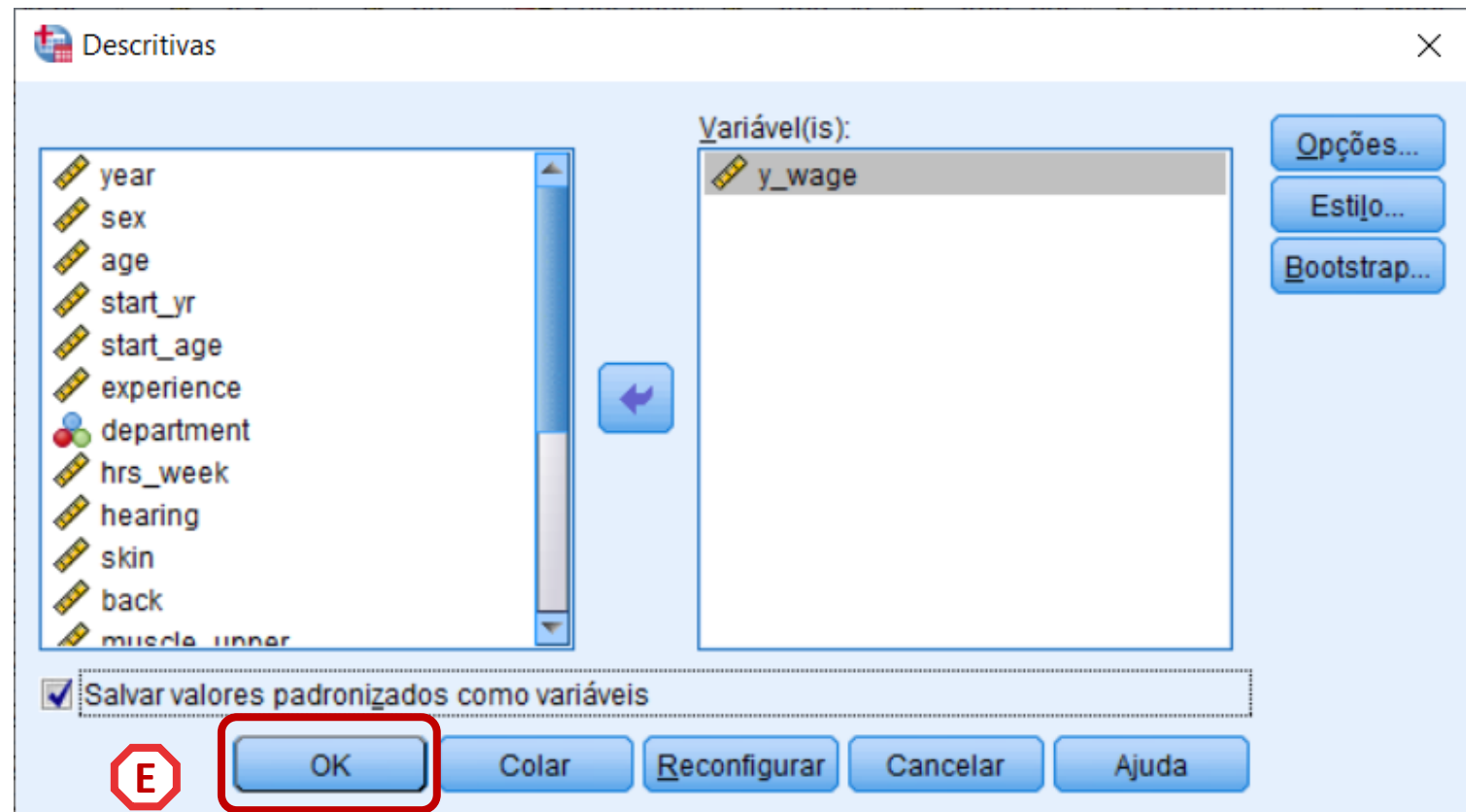
Z-Scores para identificar outliers

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Descritivas' **A**
- Selecionar a variável 'y_wage' **B**
- Colocar na caixa 'Variável(is)' **C**
- Seleccionar a opção 'Salvar valores padronizados como variáveis' (z-scores) **D**



Z-Scores para identificar outliers

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Descritivas' **A**
- Selecionar a variável 'y_wage' **B**
- Colocar na caixa 'Variável(is)' **C**
- Seleccionar a opção 'Salvar valores padronizados como variáveis' (z-scores) **D**
- Clicar 'OK' / 'Continuar' **E**



Z-Scores para identificar outliers

- No ambiente 'Visualização de Dados' podemos ver a variável com os z-scores associados



*ADGRH_BD_Aula3a.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: Zy_wage -59726761207456 Visível: 23 de 23 variáveis

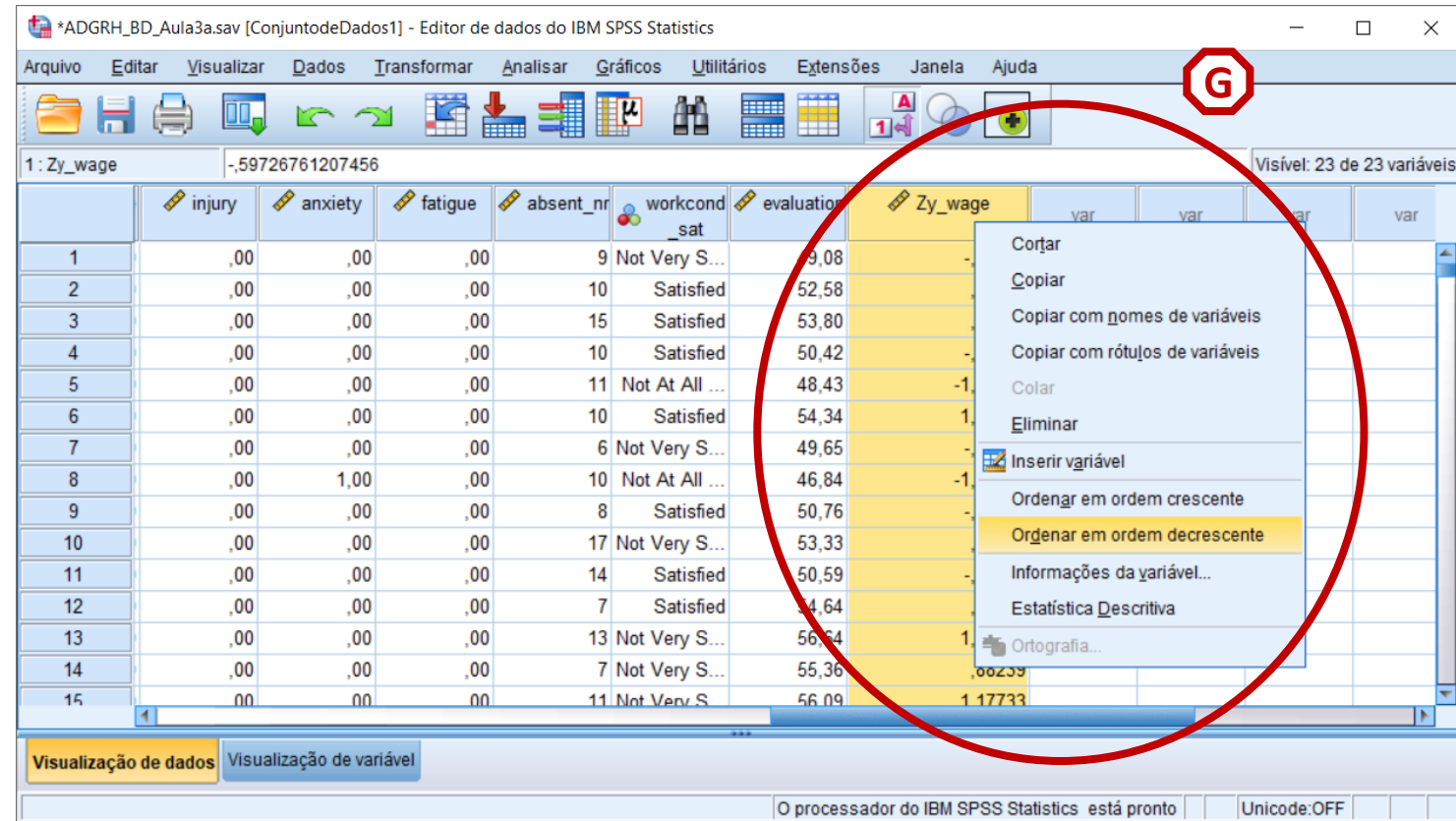
	injury	anxiety	fatigue	absent_nr	workcond_sat	evaluation	Zy_wage	var	var	var	var
1	,00	,00	,00	9	Not Very S...	49,08	-,59727				
2	,00	,00	,00	10	Satisfied	52,38	,40815				
3	,00	,00	,00	15	Satisfied	53,00	,71027				
4	,00	,00	,00	10	Satisfied	50,42	-,71326				
5	,00	,00	,00	11	Not At All ...	48,43	-1,55112				
6	,00	,00	,00	10	Satisfied	54,34	1,00263				
7	,00	,00	,00	6	Not Very S...	49,05	-,82585				
8	,00	1,00	,00	10	Not At All ...	46,04	-1,88661				
9	,00	,00	,00	8	Satisfied	50,76	-,82315				
10	,00	,00	,00	17	Not Very S...	53,33	,73204				
11	,00	,00	,00	14	Satisfied	50,39	-,60733				
12	,00	,00	,00	7	Satisfied	54,04	,49816				
13	,00	,00	,00	13	Not Very S...	56,04	1,03958				
14	,00	,00	,00	7	Not Very S...	55,36	,88239				
15	,00	,00	,00	11	Not Very S...	56,00	1,17733				

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF

Z-Scores para identificar outliers

- No ambiente 'Visualização de Dados' podemos ver a variável com os z-scores associados
- Para melhor podermos identificar os valores extremos, convém ordenar a variável por ordem decrescente
- Para isso carregamos com o botão direito do rato sobre o identificador da variável e selecionamos a opção 'Ordenar em ordem decrescente'



Z-Scores para identificar outliers

- São considerados 'valores extremos' observações em que o valor do z-score:
- É > 3.28



*ADGRH_BD_Aula3a.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

5 : Zy_wage 17,18821734656477 Visível: 23 de 23 variáveis

	injury	anxiety	fatigue	absent_nr	workcond_sat	evaluation	Zy_wage	var	var	var	var
1	,00	,00	,00	7	Satisfied	61,56	17,91212				
2	,00	,00	,00	7	Very satisfi...	60,73	17,20772				
3	,00	,00	,00	11	Not Very S...	61,56	17,25841				
4	,00	,00	,00	15	Satisfied	60,71	17,19750				
5	,00	1,00	1,00	13	Not At All ...	60,72	17,18822				
6	,00	,00	,00	10	Very satisfi...	60,70	2,69948				
7	,00	,00	,00	8	Not At All ...	60,57	2,65896				
8	,00	,00	,00	14	Very satisfi...	61,58	2,61417				
9	,00	,00	,00	18	Not Very S...	60,54	2,34006				
10	,00	,00	,00	13	Satisfied	59,87	2,51893				
11	,00	,00	,00	9	Very satisfi...	59,84	2,50596				
12	1,00	,00	1,00	6	Very satisfi...	59,93	2,47247				
13	,00	,00	,00	11	Very satisfi...	61,21	2,46753				
14	,00	,00	,00	8	Very satisfi...	60,28	2,45139				
15	,00	,00	,00	15	Not Very S...	60,76	2,43480				

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode: OFF



Z-Scores para identificar outliers

- São considerados 'valores extremos' observações em que o valor do z-score:

- > 3.28

ou

- $< - 3.28$

*Porquê este valor? Numa distribuição típica, a maioria dos dados (cerca de 99.7%) estará dentro de ± 3 desvios padrão da média. Apenas 0.3% dos dados está além de ± 3 desvios padrão.



*ADGRH_BD_Aula3a.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

5 : Zy_wage 17,18821734656477 Visível: 23 de 23 variáveis

	injury	anxiety	fatigue	absent_nr	workcond_sat	evaluation	Zy_wage	var	var	var	var
1	,00	,00	,00	7	Satisfied	61,56	17,91212				
2	,00	,00	,00	7	Very satisfi...	60,73	17,20772				
3	,00	,00	,00	11	Not Very S...	61,56	17,25841				
4	,00	,00	,00	15	Satisfied	60,71	17,19750				
5	,00	1,00	1,00	13	Not At All ...	60,72	17,18822				
6	,00	,00	,00	10	Very satisfi...	60,70	2,69948				
7	,00	,00	,00	8	Not At All ...	60,57	2,65896				
8	,00	,00	,00	14	Very satisfi...	61,58	2,61417				
9	,00	,00	,00	18	Not Very S...	60,54	2,34006				
10	,00	,00	,00	13	Satisfied	59,87	2,51893				
11	,00	,00	,00	9	Very satisfi...	59,84	2,50596				
12	1,00	,00	1,00	6	Very satisfi...	59,93	2,47247				
13	,00	,00	,00	11	Very satisfi...	61,21	2,46753				
14	,00	,00	,00	8	Very satisfi...	60,28	2,45139				
15	,00	,00	,00	15	Not Very S...	60,76	2,43480				

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode: OFF

Limpar a Base de Dados

Recodificar valores extremos

Objetivo: recodificar valores acima de '40700'
como valores omissos

-

Z-Scores para identificar outliers

- Uma vez identificadas as observações com valores extremos, devemos olhar para os valores nas outras variáveis para avaliar se este valor extremo é natural ou resulta de um erro.

*ADGRH_BD_Aula3a.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

5 : year 2018,00 Visível: 23 de 23 variáveis

	y_wage	departme nt	hrs_week	hearing	skin	back	muscle_u pper	muscle_l ower	headache	injury	anxiety	fatigue	absent_nr	workcond _sat	evaluation	Zy_wage	var
1	126283,72	Financial	38,05	,00	,00	,00	,00	,00	,00	,00	,00	,00	7	Satisfied	61,56	17,91212	
2	122883,46	IT	42,01	,00	,00	,00	,00	,00	,00	,00	,00	,00	7	Very satisfi...	60,73	17,30772	
3	122606,01	Marketing	39,96	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Not Very S...	61,38	17,25841	
4	122263,37	Marketing	38,53	,00	,00	,00	,00	,00	,00	,00	,00	,00	15	Satisfied	60,71	17,19750	
5	122211,14	Admin	40,22	,00	,00	,00	,00	,00	1,00	,00	1,00	1,00	13	Not At All ...	60,72	17,18822	
6	40699,97	Financial	39,26	,00	,00	,00	,00	,00	,00	,00	,00	,00	10	Very satisfi...	60,70	2,69948	
7	40472,01	Production	38,84	,00	,00	,00	,00	,00	,00	,00	,00	,00	8	Not At All ...	60,57	2,65896	
8	40220,03	Marketing	39,49	,00	,00	,00	,00	,00	,00	,00	,00	,00	14	Very satisfi...	61,58	2,61417	
9	39803,13	Production	40,23	,00	,00	,00	,00	,00	,00	,00	,00	,00	18	Not Very S...	60,54	2,54006	
10	39684,23	Sales	39,80	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Satisfied	59,87	2,51893	
11	39611,27	Financial	40,34	,00	,00	,00	,00	,00	,00	,00	,00	,00	9	Very satisfi...	59,84	2,50596	
12	39422,85	Production	39,10	1,00	1,00	1,00	1,00	1,00	,00	1,00	,00	1,00	6	Very satisfi...	59,93	2,47247	
13	39395,07	Sales	39,53	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Very satisfi...	61,21	2,46753	
14	39304,26	Audit	39,79	,00	,00	,00	,00	,00	,00	,00	,00	,00	8	Very satisfi...	60,28	2,45139	
15	39210,94	Admin	40,14	,00	,00	,00	,00	,00	,00	,00	,00	,00	15	Not Very S...	60,76	2,43480	
16	39035,41	Production	40,77	,00	,00	,00	,00	,00	,00	,00	,00	,00	6	Not Very S...	60,12	2,40360	
17	38710,39	IT	39,02	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Not Very S...	59,33	2,34582	
18	38592,13	Sales	41,65	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Not Very S...	59,13	2,32480	
19	38529,44	Logistics	38,75	,00	,00	,00	,00	,00	,00	,00	,00	,00	7	Satisfied	59,88	2,31366	
20	38471,41	Marketing	41,34	,00	,00	,00	,00	,00	,00	,00	,00	,00	8	Not Very S...	59,10	2,30335	
21	38352,30	Sales	40,13	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Satisfied	59,64	2,28217	
22	38319,75	Sales	39,79	,00	,00	,00	,00	,00	,00	,00	,00	,00	9	Satisfied	58,95	2,27639	
23	38271,16	Production	40,71	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Not At All ...	60,13	2,26775	
24	38031,28	Financial	39,04	,00	,00	,00	,00	,00	,00	,00	,00	,00	12	Very satisfi...	58,83	2,22511	
25	37953,69	Admin	39,91	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Not Very S...	59,42	2,21132	
26	37815,63	Marketing	39,96	,00	,00	,00	,00	,00	,00	,00	,00	,00	16	Satisfied	58,78	2,18678	
27	37427,57	Financial	40,18	,00	,00	,00	,00	,00	,00	,00	,00	,00	6	Satisfied	58,37	2,11780	

Visualização de dados Visualização de variável

Desfazer uma ação do usuário O processador do IBM SPSS Statistics está pronto Unicode:OFF

93

Recodificar como Valor Omisso

- Se olharmos com atenção...

*ADGRH_BD_Aula3a.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

5 : year 2018,00 Visível: 23 de 23 variáveis

	experiencia	salario	departamento	hrs_week	hearing	skin	back	muscle_upper	muscle_lower	headache	injury	anxiety	fatigue	absent_nr	workcond_sat	evaluation	Zy_wage	va
1	17,00	126283,72	Financial	38,05	,00	,00	,00	,00	,00	,00	,00	,00	,00	7	Satisfied	61,56	17,91212	
2	20,00	122883,46	IT	42,01	,00	,00	,00	,00	,00	,00	,00	,00	,00	7	Very satisfi...	60,73	17,30772	
3	15,00	122606,01	Marketing	39,96	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Not Very S...	61,38	17,25841	
4	15,00	122263,37	Marketing	38,53	,00	,00	,00	,00	,00	,00	,00	,00	,00	15	Satisfied	60,71	17,19750	
5	14,00	122211,14	Admin	40,22	,00	,00	,00	,00	,00	1,00	,00	1,00	1,00	13	Not At All ...	60,72	17,18822	
6	14,00	40699,97	Financial	39,26	,00	,00	,00	,00	,00	,00	,00	,00	,00	10	Very satisfi...	60,70	2,69948	
7	13,00	40472,01	Production	38,84	,00	,00	,00	,00	,00	,00	,00	,00	,00	8	Not At All ...	60,57	2,65896	
8	13,00	40220,03	Marketing	39,49	,00	,00	,00	,00	,00	,00	,00	,00	,00	14	Very satisfi...	61,58	2,61417	
9	20,00	39803,13	Production	40,23	,00	,00	,00	,00	,00	,00	,00	,00	,00	18	Not Very S...	60,54	2,54006	
10	20,00	39684,23	Sales	39,80	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Satisfied	59,87	2,51893	
11	19,00	39611,27	Financial	40,34	,00	,00	,00	,00	,00	,00	,00	,00	,00	9	Very satisfi...	59,84	2,50596	
12	10,00	39422,85	Production	39,10	1,00	1,00	1,00	1,00	1,00	,00	1,00	,00	1,00	6	Very satisfi...	59,93	2,47247	
13	18,00	39395,07	Sales	39,53	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Very satisfi...	61,21	2,46753	
14	17,00	39304,26	Audit	39,79	,00	,00	,00	,00	,00	,00	,00	,00	,00	8	Very satisfi...	60,28	2,45139	
15	19,00	39210,94	Admin	40,14	,00	,00	,00	,00	,00	,00	,00	,00	,00	15	Not Very S...	60,76	2,43480	
16	16,00	39035,41	Production	40,77	,00	,00	,00	,00	,00	,00	,00	,00	,00	6	Not Very S...	60,12	2,40360	
17	15,00	38710,39	IT	39,02	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Not Very S...	59,33	2,34582	
18	20,00	38592,13	Sales	41,65	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Not Very S...	59,13	2,32480	
19	12,00	38529,44	Logistics	38,75	,00	,00	,00	,00	,00	,00	,00	,00	,00	7	Satisfied	59,88	2,31366	
20	18,00	38471,41	Marketing	41,34	,00	,00	,00	,00	,00	,00	,00	,00	,00	8	Not Very S...	59,10	2,30335	
21	17,00	38352,30	Sales	40,13	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Satisfied	59,64	2,28217	
22	20,00	38319,75	Sales	39,79	,00	,00	,00	,00	,00	,00	,00	,00	,00	9	Satisfied	58,95	2,27639	
23	19,00	38271,16	Production	40,71	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Not At All ...	60,13	2,26775	
24	17,00	38031,28	Financial	39,04	,00	,00	,00	,00	,00	,00	,00	,00	,00	12	Very satisfi...	58,83	2,22511	
25	15,00	37953,69	Admin	39,91	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Not Very S...	59,42	2,21132	
26	13,00	37815,63	Marketing	39,96	,00	,00	,00	,00	,00	,00	,00	,00	,00	16	Satisfied	58,78	2,18678	
27	19,00	37427,57	Financial	40,18	,00	,00	,00	,00	,00	,00	,00	,00	,00	6	Satisfied	58,37	2,11780	

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF

Recodificar como Valor Omisso

- Se olharmos com atenção...
- Podemos identificar qual o valor da variável 'y_wage' a partir do qual os valores são considerados extremos
- Neste caso, todos os valores acima de 40699,97 são considerados valores extremos



*ADGRH_BD_Aula3a.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

5 : year 2018,00 Visível: 23 de 23 variáveis

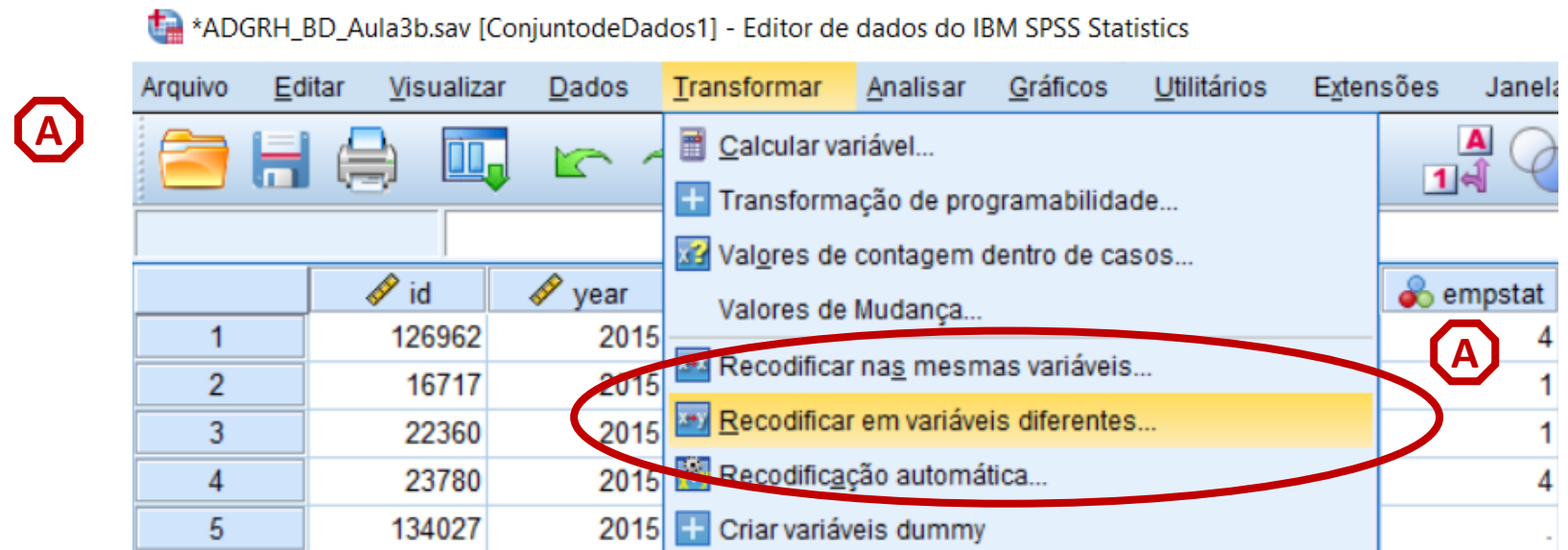
	experiencia	y_wage	departamento	hrs_week	hearing	skin	back	muscle_upper	muscle_lower	headache	injury	anxiety	fatigue	absent_nr	workcond_sat	evaluation	Zy_wage	va
1	17,00	126283,72	Financial	38,05	,00	,00	,00	,00	,00	,00	,00	,00	,00	7	Satisfied	61,56	17,91212	
2	20,00	122883,46	IT	42,01	,00	,00	,00	,00	,00	,00	,00	,00	,00	7	Very satisfi...	60,73	17,30772	
3	15,00	122606,01	Marketing	39,96	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Not Very S...	61,38	17,25841	
4	15,00	122263,37	Marketing	38,53	,00	,00	,00	,00	,00	,00	,00	,00	,00	15	Satisfied	60,71	17,19750	
5	14,00	122211,14	Admin	40,22	,00	,00	,00	,00	,00	1,00	,00	1,00	1,00	13	Not At All ...	60,72	17,18822	
6	14,00	40699,97	Financial	39,26	,00	,00	,00	,00	,00	,00	,00	,00	,00	10	Very satisfi...	60,70	2,69948	
7	13,00	40472,01	Production	38,84	,00	,00	,00	,00	,00	,00	,00	,00	,00	8	Not At All ...	60,57	2,65896	
8	13,00	40220,03	Marketing	39,49	,00	,00	,00	,00	,00	,00	,00	,00	,00	14	Very satisfi...	61,56	2,61417	
9	20,00	39803,13	Production	40,23	,00	,00	,00	,00	,00	,00	,00	,00	,00	18	Not Very S...	60,70	2,54006	
10	20,00	39684,23	Sales	39,80	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Satisfied	59,84	2,51893	
11	19,00	39611,27	Financial	40,34	,00	,00	,00	,00	,00	,00	,00	,00	,00	9	Very satisfi...	59,84	2,50596	
12	10,00	39422,85	Production	39,10	1,00	1,00	1,00	1,00	1,00	,00	1,00	,00	1,00	6	Very satisfi...	59,93	2,47247	
13	18,00	39395,07	Sales	39,53	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Very satisfi...	61,21	2,46753	
14	17,00	39304,26	Audit	39,79	,00	,00	,00	,00	,00	,00	,00	,00	,00	8	Very satisfi...	60,28	2,45139	
15	19,00	39210,94	Admin	40,14	,00	,00	,00	,00	,00	,00	,00	,00	,00	15	Not Very S...	60,76	2,43480	
16	16,00	39035,41	Production	40,77	,00	,00	,00	,00	,00	,00	,00	,00	,00	6	Not Very S...	60,12	2,40360	
17	15,00	38710,39	IT	39,02	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Not Very S...	59,33	2,34582	
18	20,00	38592,13	Sales	41,65	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Not Very S...	59,13	2,32480	
19	12,00	38529,44	Logistics	38,75	,00	,00	,00	,00	,00	,00	,00	,00	,00	7	Satisfied	59,88	2,31366	
20	18,00	38471,41	Marketing	41,34	,00	,00	,00	,00	,00	,00	,00	,00	,00	8	Not Very S...	59,10	2,30335	
21	17,00	38352,30	Sales	40,13	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Satisfied	59,64	2,28217	
22	20,00	38319,75	Sales	39,79	,00	,00	,00	,00	,00	,00	,00	,00	,00	9	Satisfied	58,95	2,27639	
23	19,00	38271,16	Production	40,71	,00	,00	,00	,00	,00	,00	,00	,00	,00	11	Not At All ...	60,13	2,26775	
24	17,00	38031,28	Financial	39,04	,00	,00	,00	,00	,00	,00	,00	,00	,00	12	Very satisfi...	58,83	2,22511	
25	15,00	37953,69	Admin	39,91	,00	,00	,00	,00	,00	,00	,00	,00	,00	13	Not Very S...	59,42	2,21132	
26	13,00	37815,63	Marketing	39,96	,00	,00	,00	,00	,00	,00	,00	,00	,00	16	Satisfied	58,78	2,18678	
27	19,00	37427,57	Financial	40,18	,00	,00	,00	,00	,00	,00	,00	,00	,00	6	Satisfied	58,37	2,11780	

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF

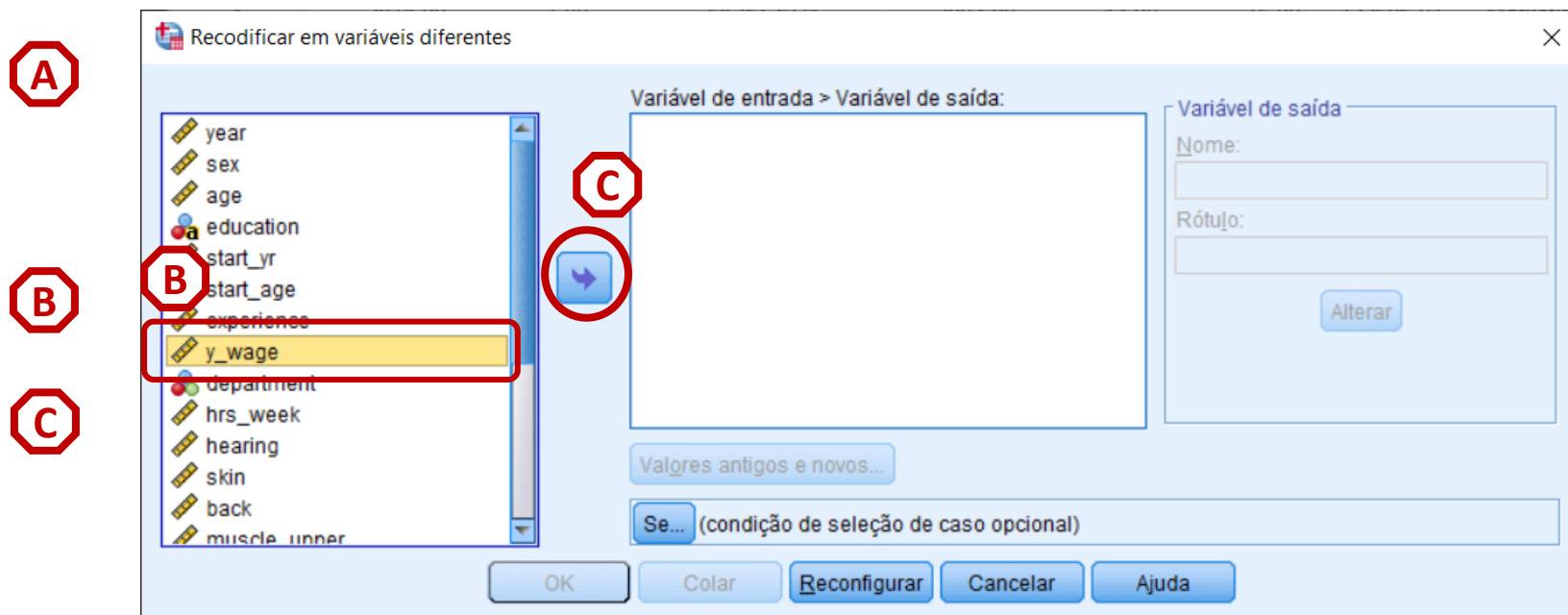
Recodificar como Valor Omisso

- Selecionar 'Transformar' / 'Recodificar em variáveis diferentes'



Recodificar como Valor Omisso

- Selecionar 'Transformar' / 'Recodificar em variáveis diferentes'
- Selecionar a variável 'y_wage'
- Colocá-la na caixa 'Variável numérica'



Recodificar como Valor Omisso

- Selecionar 'Transformar' / 'Recodificar em variáveis diferentes'
- Selecionar a variável 'y_wage'
- Colocá-la na caixa 'Variável numérica'
- Definir o nome da variável como 'y_wagen' e o rótulo como 'Wage'
- Clicar 'Alterar'
- Clicar e 'Valores antigos e novos'

A

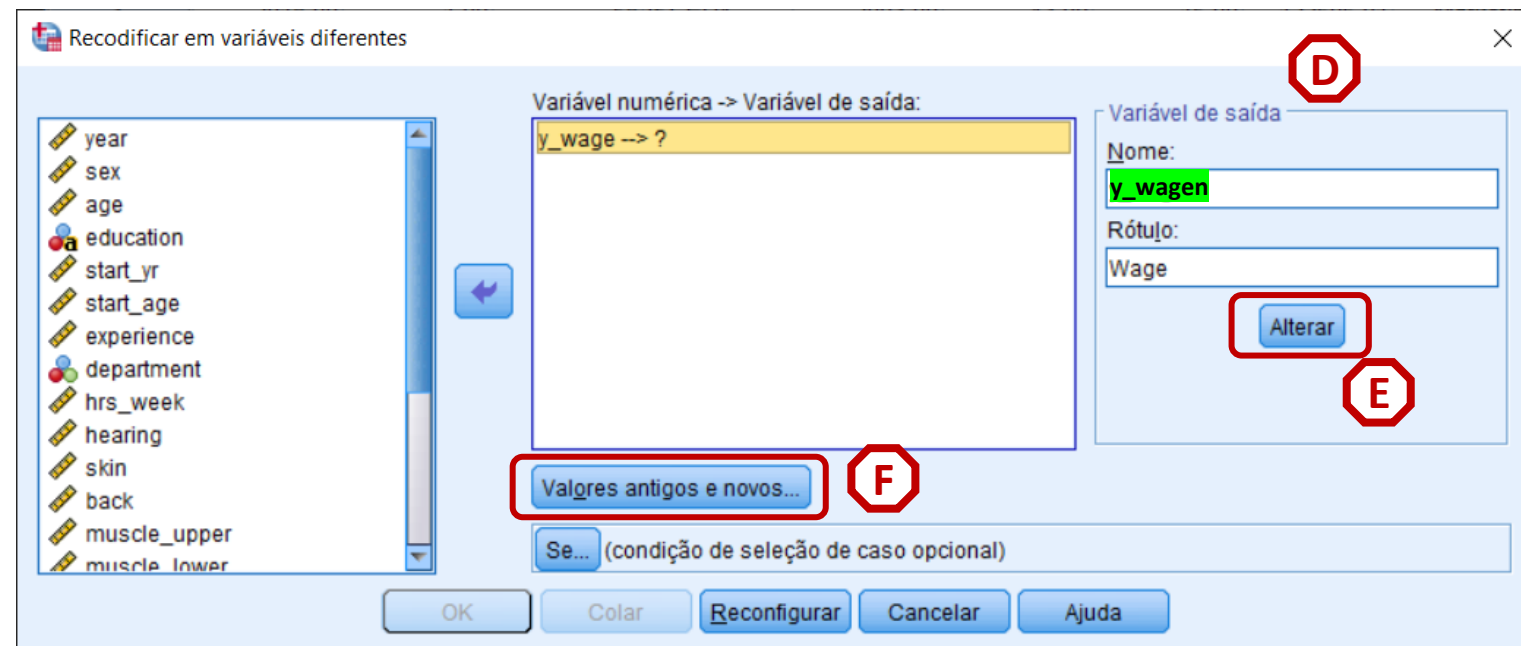
B

C

D

E

F



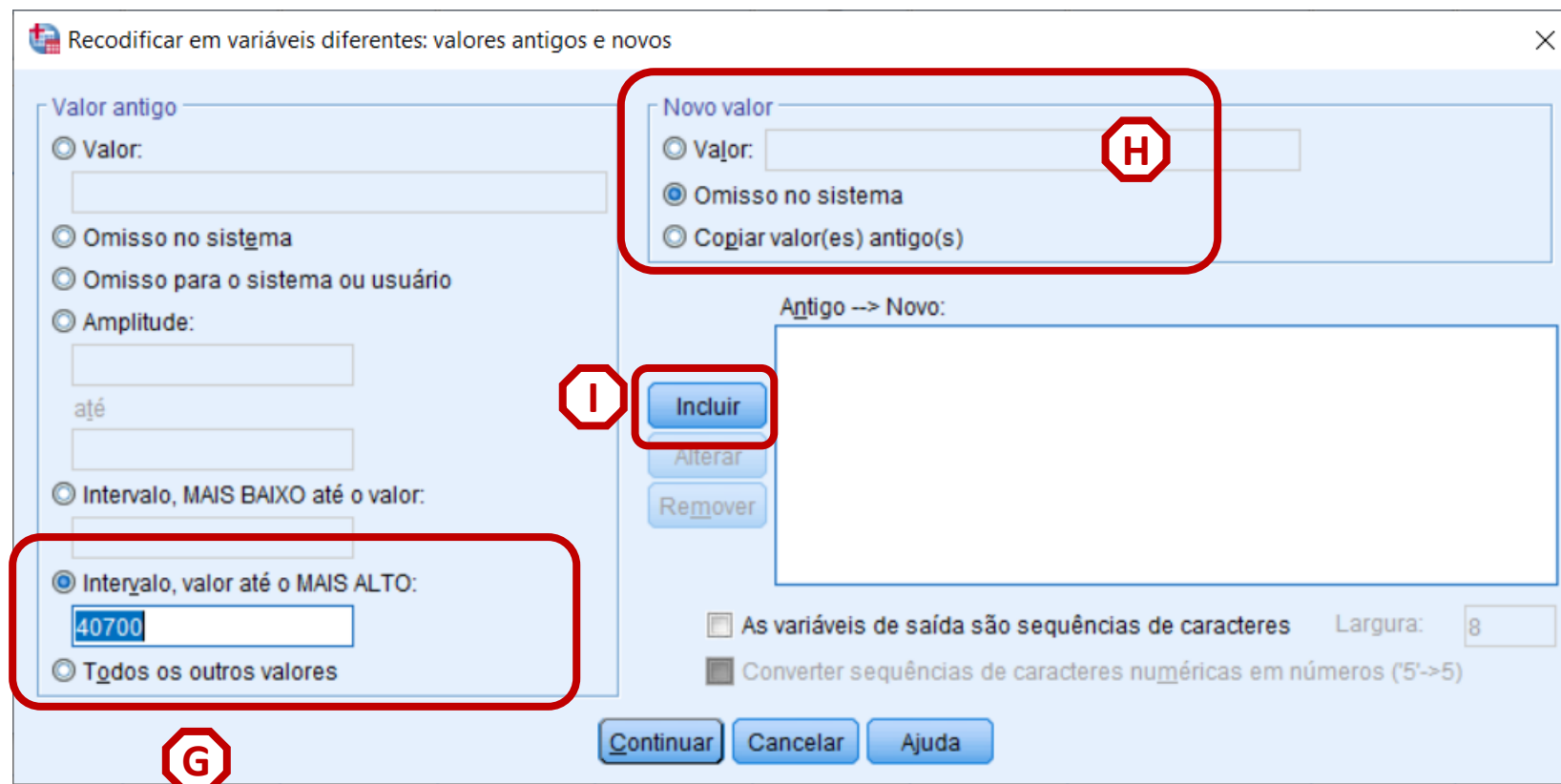
Recodificar como Valor Omisso

- Primeiro, instruímos o SPSS que todas as observações com o valor acima de '40700' ... devem ser tratadas como 'System-missing' na nova variável
- Clicar 'Incluir'

G

H

I



Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo

- ☐ Valor:
- ☐ Omisso no sistema
- ☐ Omisso para o sistema ou usuário
- ☐ Amplitude:
- ☐ Intervalo, MAIS BAIXO até o valor:
- ☒ Intervalo, valor até o MAIS ALTO: 40700
- ☐ Todos os outros valores

Novo valor

- ☐ Valor:
- ☒ Omisso no sistema
- ☐ Copiar valor(es) antigo(s)

Antigo --> Novo:

Incluir

Alterar
Remover

☐ As variáveis de saída são sequências de caracteres Largura: 8

☐ Converter sequências de caracteres numéricas em números ('5' -> 5)

Continuar Cancelar Ajuda

G

Recodificar como Valor Omisso

- Primeiro, instruímos o SPSS que todas as observações com o valor '99999999'
- ... devem ser tratadas como 'System- missing' na nova variável
- Clicar 'Incluir'



Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo

☐ Valor:

☐ Omisso no sistema

☐ Omisso para o sistema ou usuário

☐ Amplitude:

até

☐ Intervalo, MAIS BAIXO até o valor:

☒ Intervalo, valor até o MAIS ALTO:

☐ Todos os outros valores

Novo valor

☐ Valor:

☒ Omisso no sistema

☐ Copiar valor(es) antigo(s)

Antigo --> Novo:

40700 thru Highest --> SYSMIS

Incluir

Alterar

Remover

☐ As variáveis de saída são sequências de caracteres Largura: 8

☐ Converter sequências de caracteres numéricas em números ('5' -> 5)

Continuar Cancelar Ajuda

Recodificar como Valor Omisso

- Primeiro, instruímos o SPSS que todas as observações com o valor '99999999'

... devem ser tratadas como 'System- missing' na nova variável

- Clicar 'Incluir'

- Depois, definir que 'Todos os outros valores'

... deve ser copiados ('Copiar valor(es) antigo(s)') para a nova variável

- Clicar 'Incluir'

The image shows the SPSS 'Recodificar em variáveis diferentes: valores antigos e novos' (Recode into Different Variables: Old and New Values) dialog box. The window title is 'Recodificar em variáveis diferentes: valores antigos e novos'. The dialog is divided into two main sections: 'Valor antigo' (Old Value) and 'Novo valor' (New Value). In the 'Valor antigo' section, the 'Todos os outros valores' (All other values) option is selected. In the 'Novo valor' section, the 'Copiar valor(es) antigo(s)' (Copy old value(s)) option is selected. The 'Incluir' (Include) button is highlighted. The 'Antigo --> Novo:' (Old --> New:) list shows the mapping '40700 thru Highest --> SYSMIS'. At the bottom, there are checkboxes for 'As variáveis de saída são sequências de caracteres' (Output variables are character sequences) and 'Converter sequências de caracteres numéricas em números ('5'->5)' (Convert numeric character sequences to numbers ('5' -> 5)). The 'Continuar' (Continue), 'Cancelar' (Cancel), and 'Ajuda' (Help) buttons are at the bottom right. Red octagonal callouts with letters G through L and K point to specific elements: G points to the 'Valor antigo' section, H to the 'Omisso no sistema' option, I to the 'até' label, J to the 'Intervalo, MAIS BAIXO até o valor:' option, K to the 'Novo valor' section, L to the 'Incluir' button, and another J to the 'Todos os outros valores' option.

Recodificar como Valor Omisso

- Primeiro, instruímos o SPSS que todas as observações com o valor '99999999'

... devem ser tratadas como 'System- missing' na nova variável

- Clicar 'Incluir'

- Depois, definir que 'Todos os outros valores'

... deve ser copiados ('Copiar valor(es) antigo(s)') para a nova variável

- Clicar 'Incluir'

- Clicar 'Continuar' / 'OK'

G

H

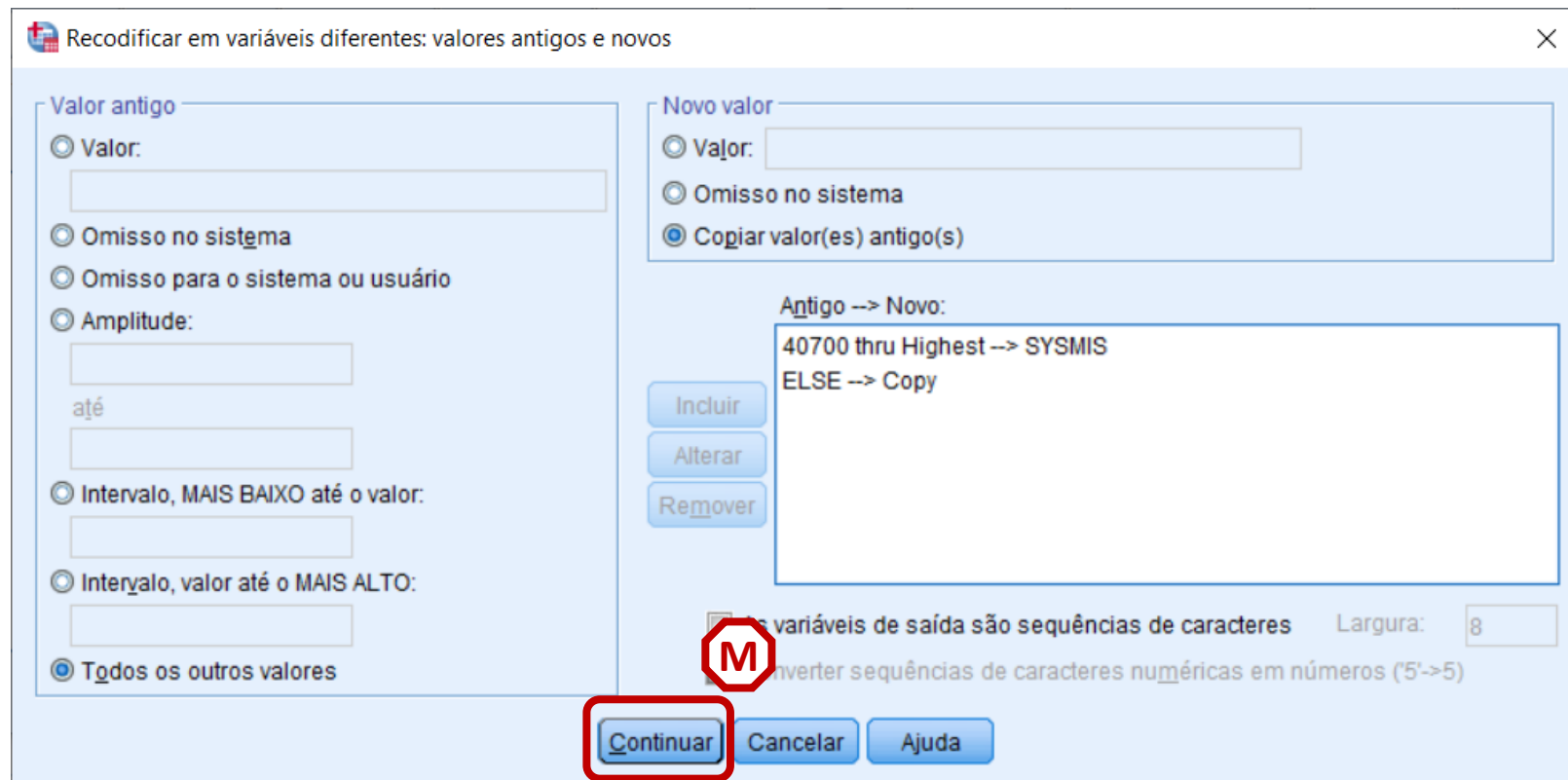
I

J

K

L

M



Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo

- ☐ Valor:
- ☐ Omisso no sistema
- ☐ Omisso para o sistema ou usuário
- ☐ Amplitude: até
- ☐ Intervalo, MAIS BAIXO até o valor:
- ☐ Intervalo, valor até o MAIS ALTO:
- ☒ Todos os outros valores

Novo valor

- ☐ Valor:
- ☐ Omisso no sistema
- ☒ Copiar valor(es) antigo(s)

Antigo --> Novo:
40700 thru Highest --> SYSMIS
ELSE --> Copy

Incluir
Alterar
Remover

☒ As variáveis de saída são sequências de caracteres Largura:
☐ Converter sequências de caracteres numéricas em números ('5' -> 5)

Continuar Cancelar Ajuda

Recodificar como Valor Omisso

- Vamos ver se a operação de recodificação foi bem sucedida
- Que outras opções falámos antes?

*ADGRH_BD_Aula3a.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : Zy_wage 17,91212371004056 Visível: 24 de 24 variáveis

	muscle_lover	headache	injury	anxiety	fatigue	absent_nr	workcond_sat	evaluation	Zy_wage	y_wagen	var
1	,00	,00	,00	,00	,00	7	Satisfied	61,56	17,91212	.	
2	,00	,00	,00	,00	,00	7	Very satisfi...	60,73	17,30772	.	
3	,00	,00	,00	,00	,00	11	Not Very S...	61,38	17,25841	.	
4	,00	,00	,00	,00	,00	15	Satisfied	60,71	17,19750	.	
5	,00	1,00	,00	1,00	1,00	13	Not At All ...	60,72	17,18822	.	
6	,00	,00	,00	,00	,00	10	Very satisfi...	60,70	2,69948	40699,97	
7	,00	,00	,00	,00	,00	8	Not At All ...	60,57	2,65896	40472,01	
8	,00	,00	,00	,00	,00	14	Very satisfi...	61,58	2,61417	40220,03	
9	,00	,00	,00	,00	,00	18	Not Very S...	60,54	2,54006	39803,13	
10	,00	,00	,00	,00	,00	13	Satisfied	59,87	2,51893	39684,23	
11	,00	,00	,00	,00	,00	9	Very satisfi...	59,84	2,50596	39611,27	
12	1,00	,00	1,00	,00	1,00	6	Very satisfi...	59,93	2,47247	39422,85	
13	,00	,00	,00	,00	,00	11	Very satisfi...	61,21	2,46753	39395,07	
14	,00	,00	,00	,00	,00	8	Very satisfi...	60,28	2,45139	39304,26	
15	,00	,00	,00	,00	,00	15	Not Very S...	60,76	2,43480	39210,94	
16	,00	,00	,00	,00	,00	6	Not Very S...	60,12	2,40360	39035,41	
17	,00	,00	,00	,00	,00	13	Not Very S...	59,33	2,34582	38710,39	
18	,00	,00	,00	,00	,00	11	Not Very S...	59,13	2,32480	38592,13	
19	,00	,00	,00	,00	,00	7	Satisfied	59,88	2,31366	38529,44	
20	,00	,00	,00	,00	,00	8	Not Very S...	59,10	2,30335	38471,41	
21	,00	,00	,00	,00	,00	13	Satisfied	59,64	2,28217	38352,30	
22	,00	,00	,00	,00	,00	9	Satisfied	58,95	2,27639	38319,75	

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF

Reportar os resultados COMPLETE O TEMPLATE

A amostra é caracterizada por 5000 participantes, $x\%$ mulheres, $x\%$ homens, com idades compreendidas entre x e x' , que totalizam uma média de x anos de idade (desvio padrão). Os participantes são trabalhadores de x departamentos, sendo os departamentos com maior representação os departamentos de x e x .

A variável avaliação varia entre x e x , sendo que os 10% dos trabalhadores com melhores pontuações tem pontuações iguais ou superiores a x e os 10% dos trabalhadores com piores pontuações apresentam resultados a iguais ou inferiores a x .

A análise exploratória dos dados identificou dados omissos (%) e dados extremos ($x\%$) na variável rendimentos. Para preparar os dados para a análise, estes casos foram substituídos pelo valor da média da variável. A variável rendimento depois desta operações varia entre x e x , sendo a distribuição de rendimentos concentrada à esquerda.

Reportar os resultados COMPLETE O TEMPLATE

Tabela 1. Caracterização da amostra: variáveis
xxxx[lista variáveis nominais e ordinais]

		N	%
Variável	Categoria		
	Categoria		
Variável	Categoria		
	Categoria		
	...		
Variável	Categoria		
	Categoria		
Total			

Tabela 2. Caracterização da amostra:
variáveis xxxx[lista variáveis intervalares
ou de razão]

	N	Min	Máximo	Média	Desvio padrão
Variável					
Variável					
Variável					
Total					

Materiais suplementares

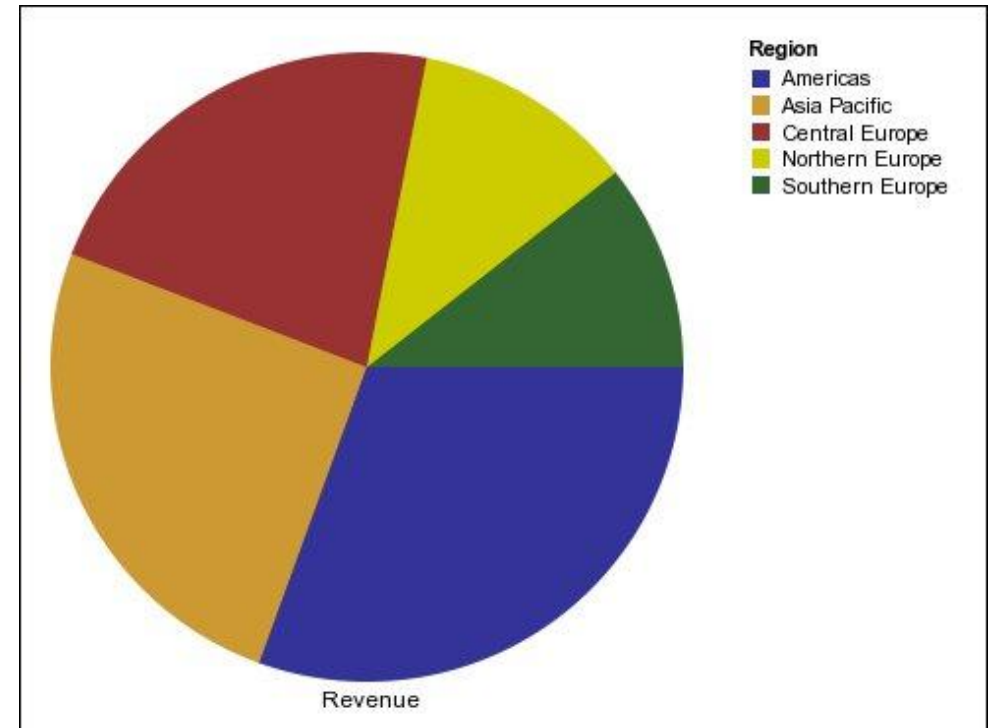
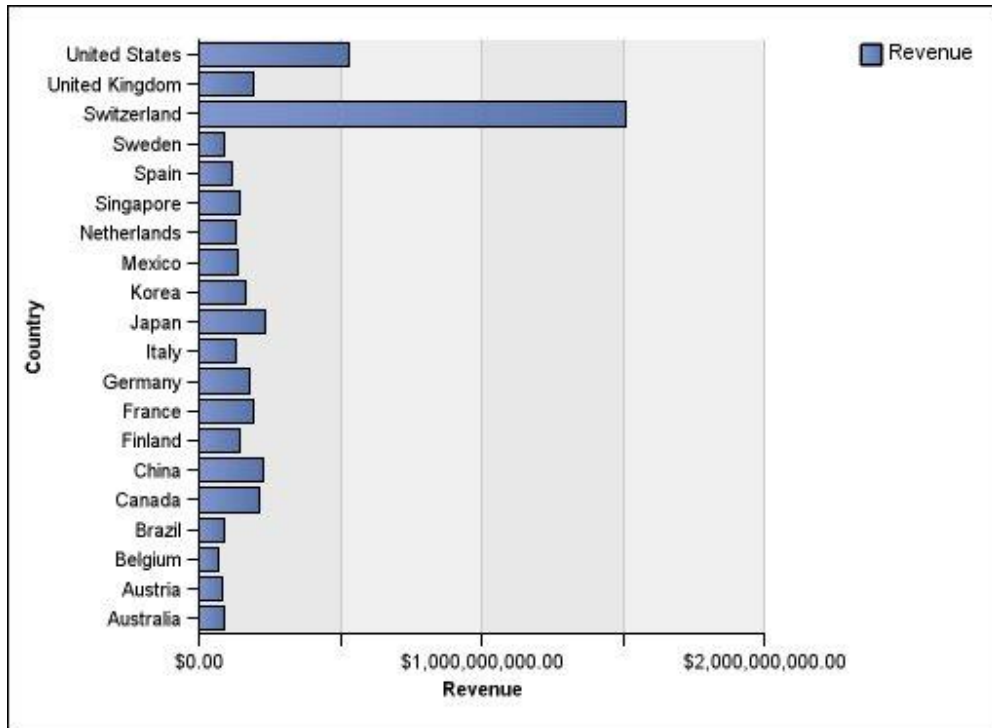
Visualização de dados com SPSS

- Qual a melhor forma de visualizar a distribuição de frequências?

Variável
Nominal
ou Ordinal

Gráfico de Barras

Gráfico Circular



- Qual a melhor forma de visualizar a distribuição de frequências?

Variável
Intervalar
ou Razão

Histograma

Histograma,
com Curva de Distribuição

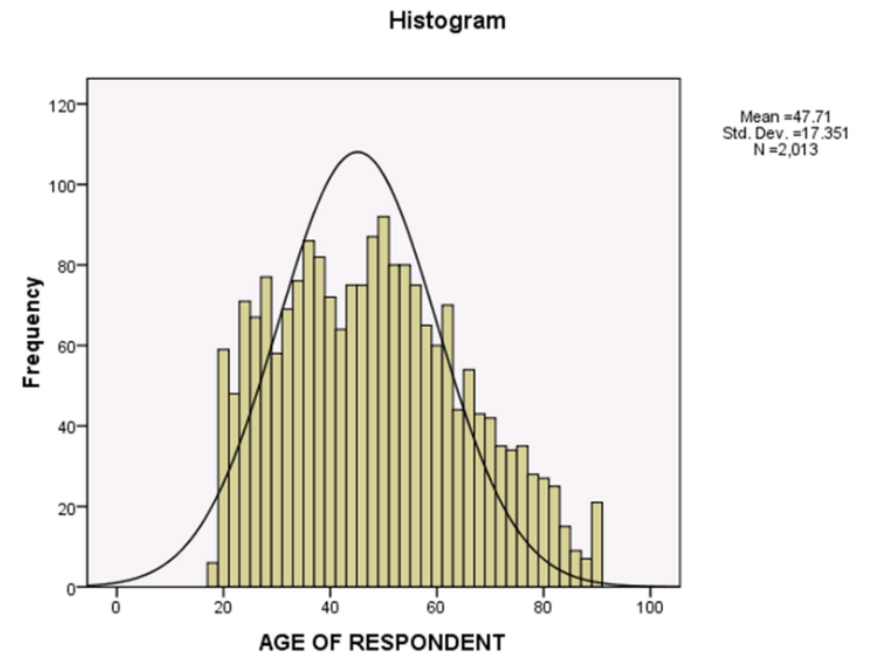
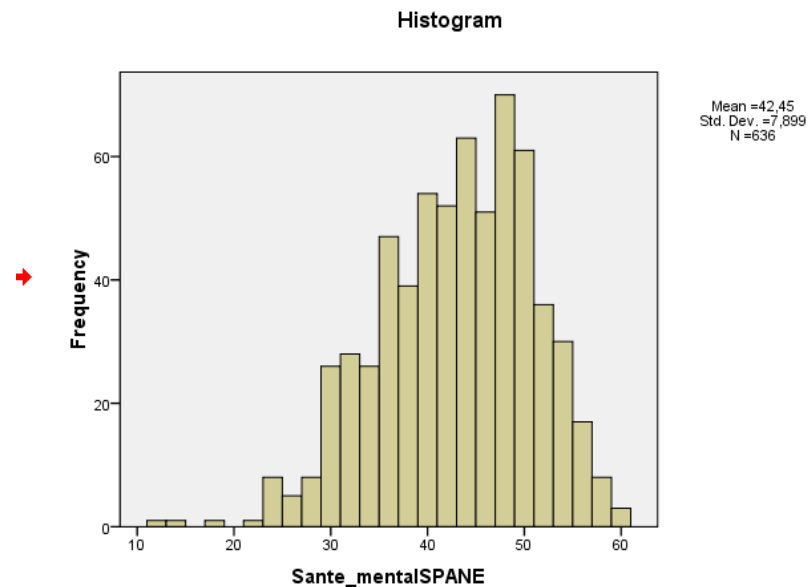


Gráfico de Barras

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'Não mostrar este diálogo novamente'
- Selecionar 'OK'

A

B

C

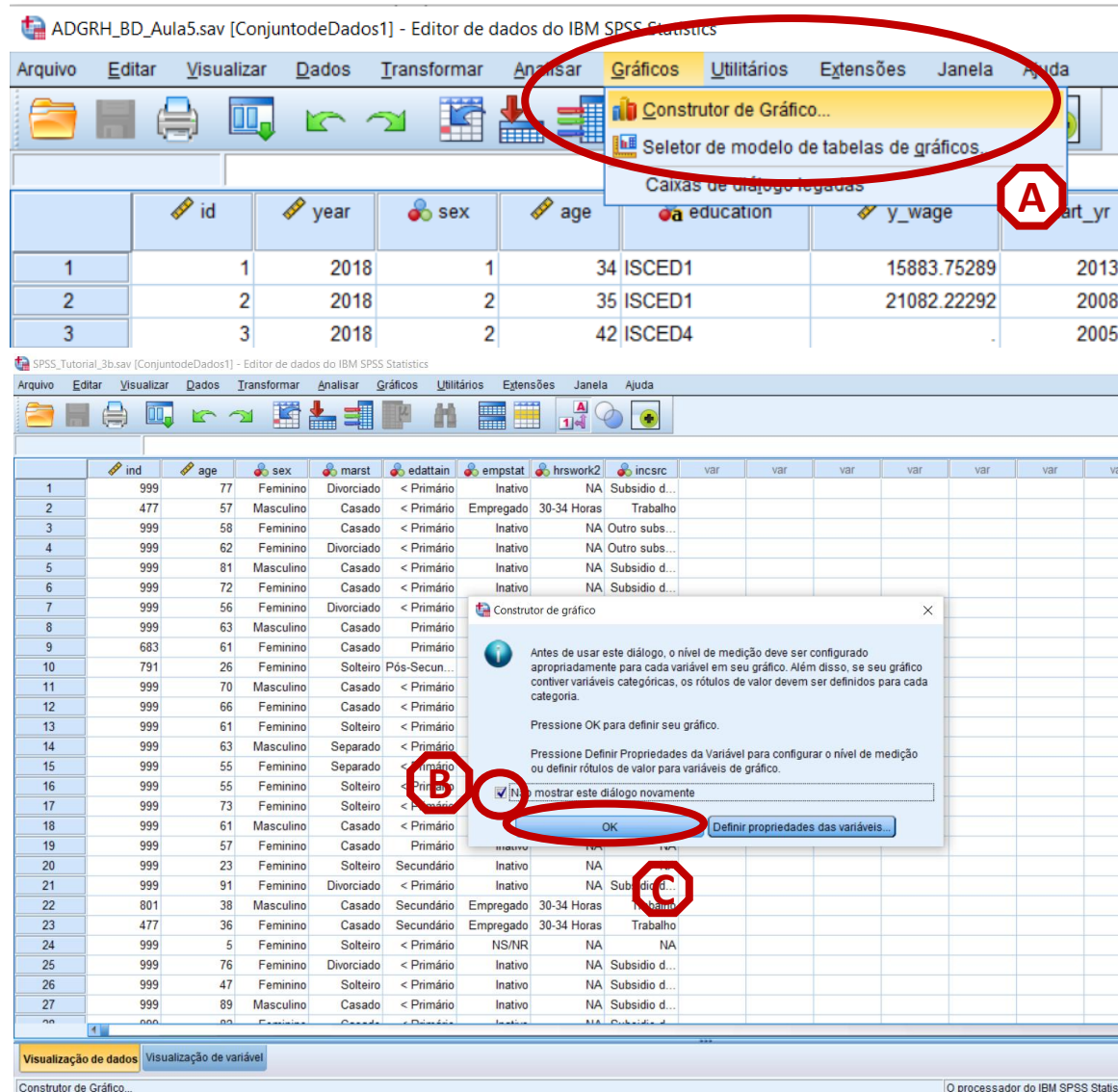


Gráfico de Barras

- Selecione 'Barras'
- Selecionar (com duplo-clique) o Gráfico de Barras (simples)

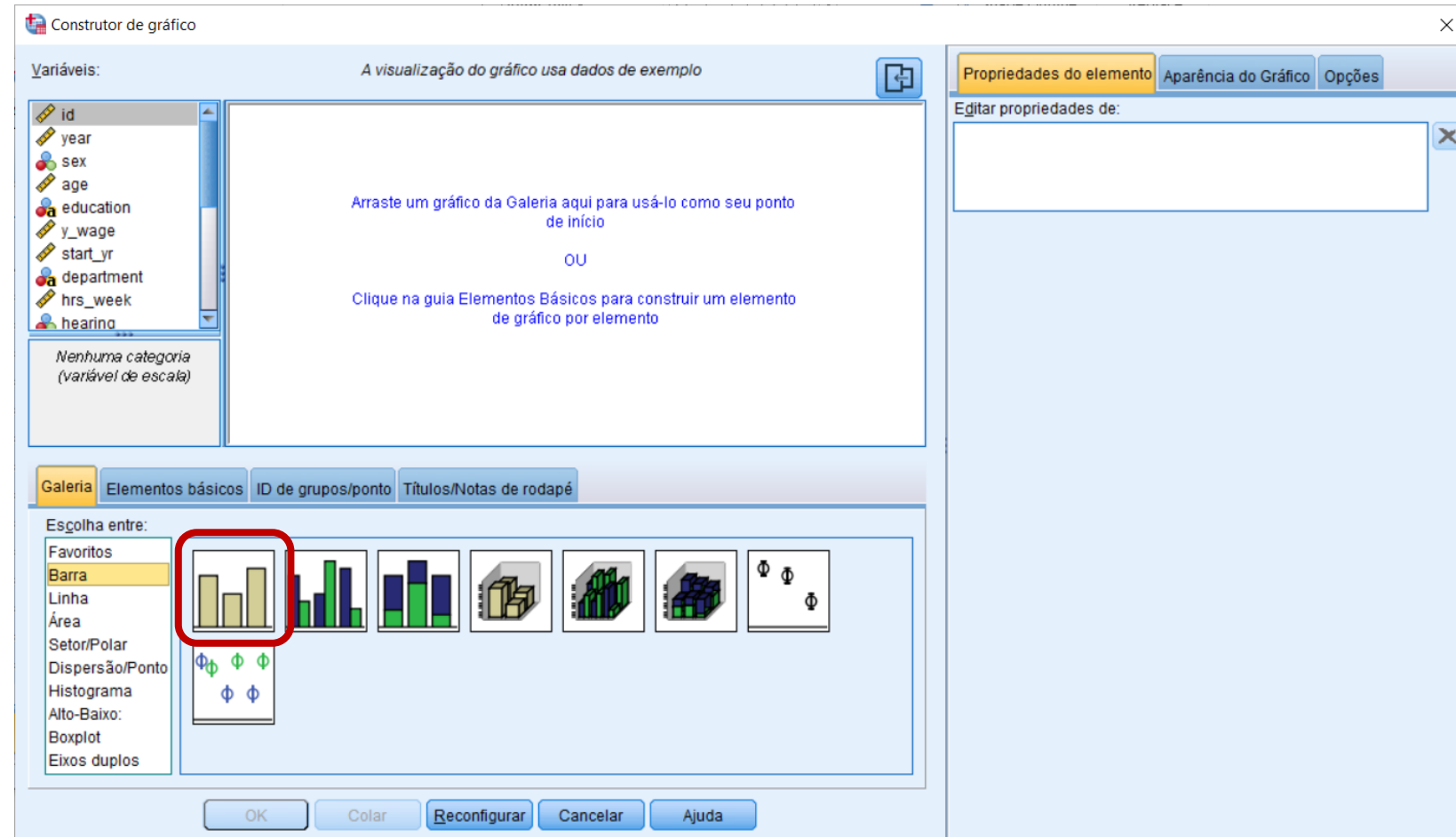


Gráfico de Barras

- Selecione 'Barras'
- Selecionar (com duplo-clique) o Gráfico de Barras (simples)
- Selecionar a variável 'department'
- Colocar a variável 'department' no 'Eixo X'

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- id
- year
- sex
- age
- education
- y_wage
- start_yr
- department
- hrs_week
- hearing
- Categoria 1
- Categoria 2

Barra Simples Contagem

Eixo Y?

Eixo X?

Galeria

Elementos básicos

ID de grupos/pontos

Notas de rodapé

Escolha entre:

- Favoritos
- Barra
- Linha
- Área
- Setor/Polar
- Dispersão/Ponto
- Histograma
- Alto-Baixo:
- Boxplot
- Eixos duplos

OK

Colar

Reconfigurar

Cancelar

Ajuda

Propriedades do elemento

Aparência do Gráfico

Opções

Editar propriedades de:

Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

Título 1

Estatísticas

Variável:

Estatística:

Contagem

Configurar parâmetros...

☒ Exibir barra de erros

Representação de Barras de Erros

☒ Intervalos de confiança

Nível (%): 95

☐ Erro padrão

Multiplicador: 2

☐ Desvio padrão

Multiplicador: 2

Estilo de barra:

Barra

Gráfico de Barras

- Selecione 'Barras'
- Selecionar (com duplo-clique) o Gráfico de Barras (simples)
- Selecionar a variável 'department'
- Colocar a variável 'department' no 'Eixo X'
- Selecionar 'OK'

D

E

F

G

H

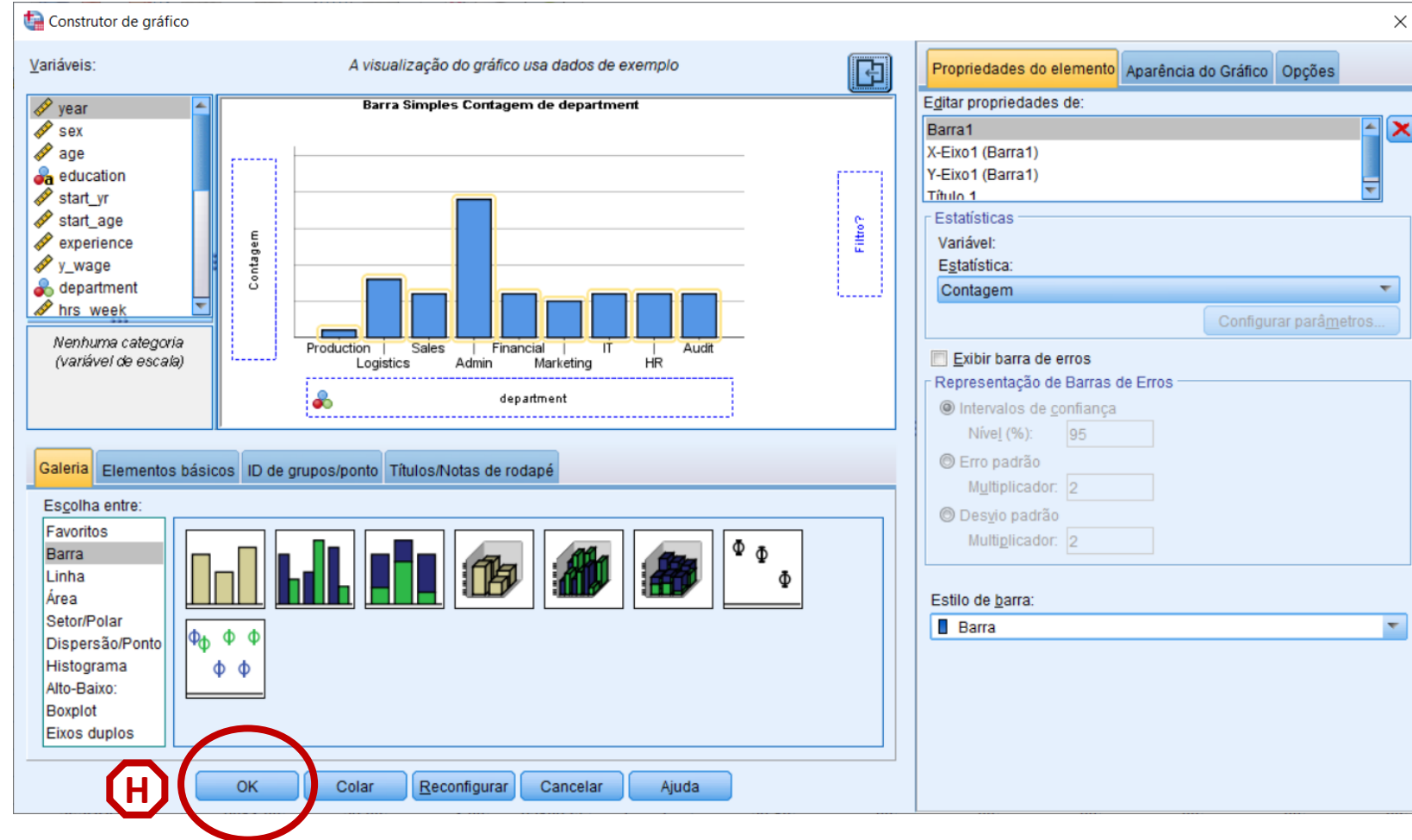


Gráfico de Barras

- O gráfico é publicado no 'Visualizador de Resultados'

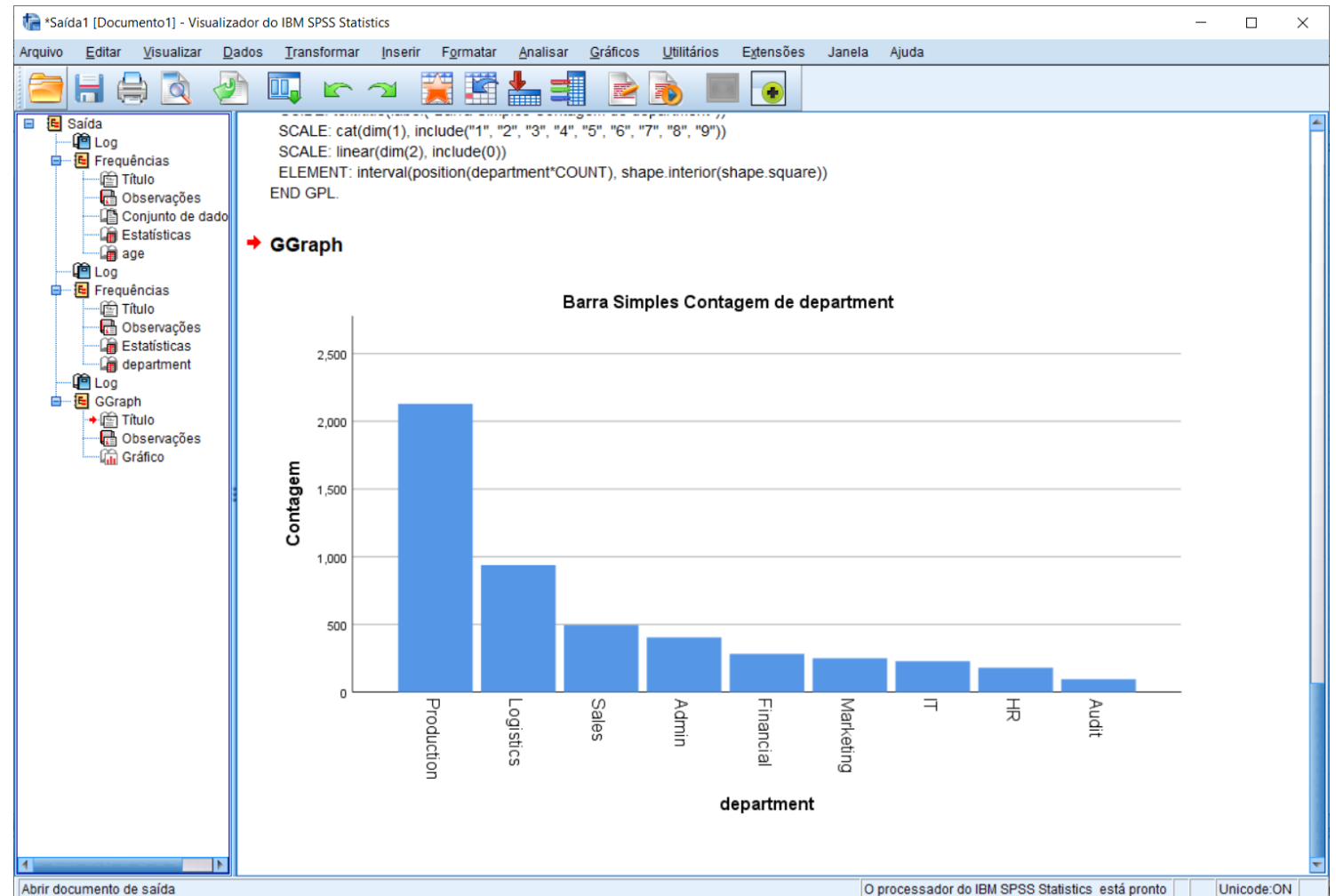


Gráfico de Barras

- O gráfico é publicado no 'Visualizador de Resultados'
- Para editarmos o gráfico, fazer duplo-clique em cima do mesmo

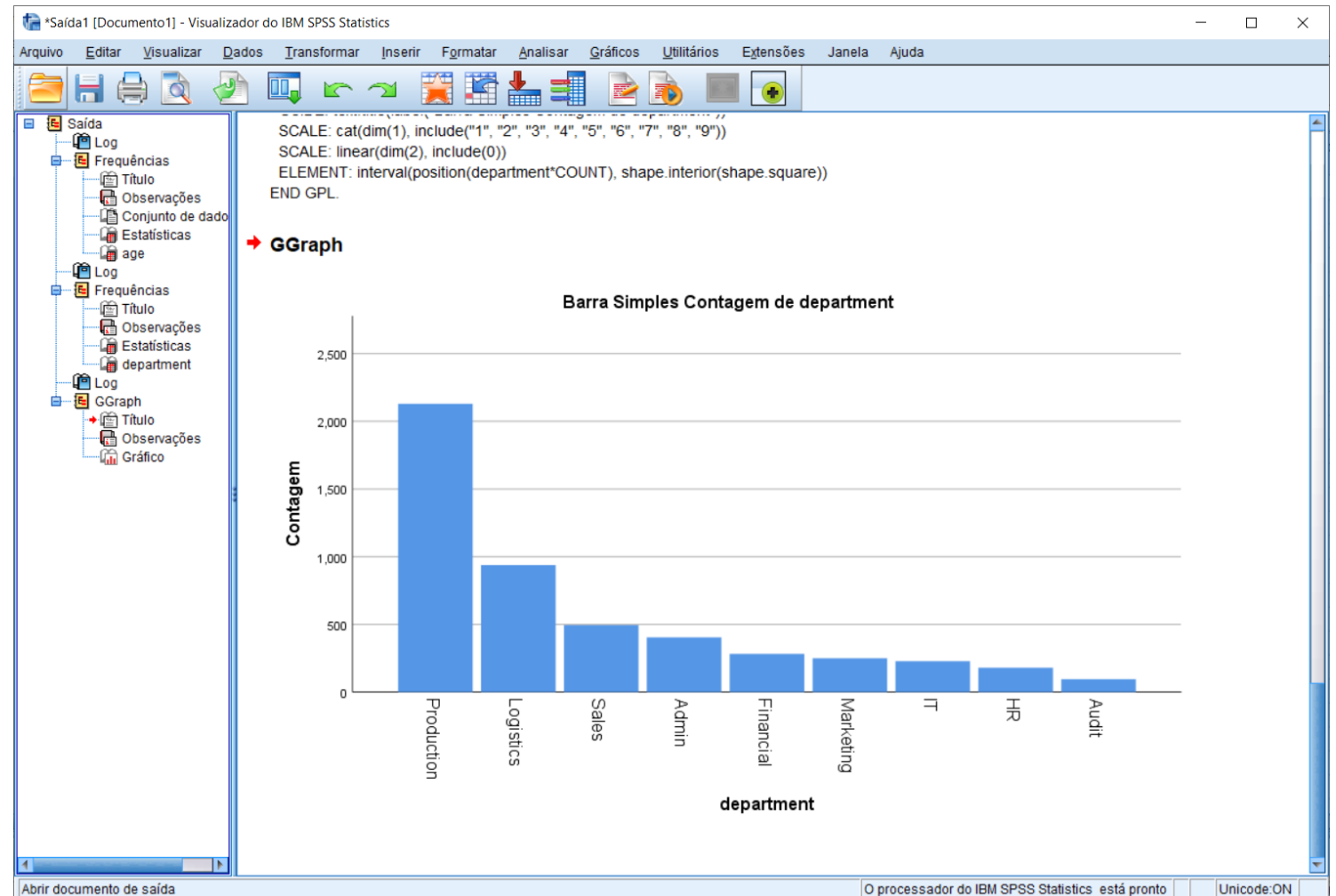


Gráfico de Barras

- O gráfico é publicado no 'Visualizador de Resultados'
- Para editarmos o gráfico, fazer duplo-clique em cima do mesmo
- Isso irá abrir o 'Editor de Gráficos'

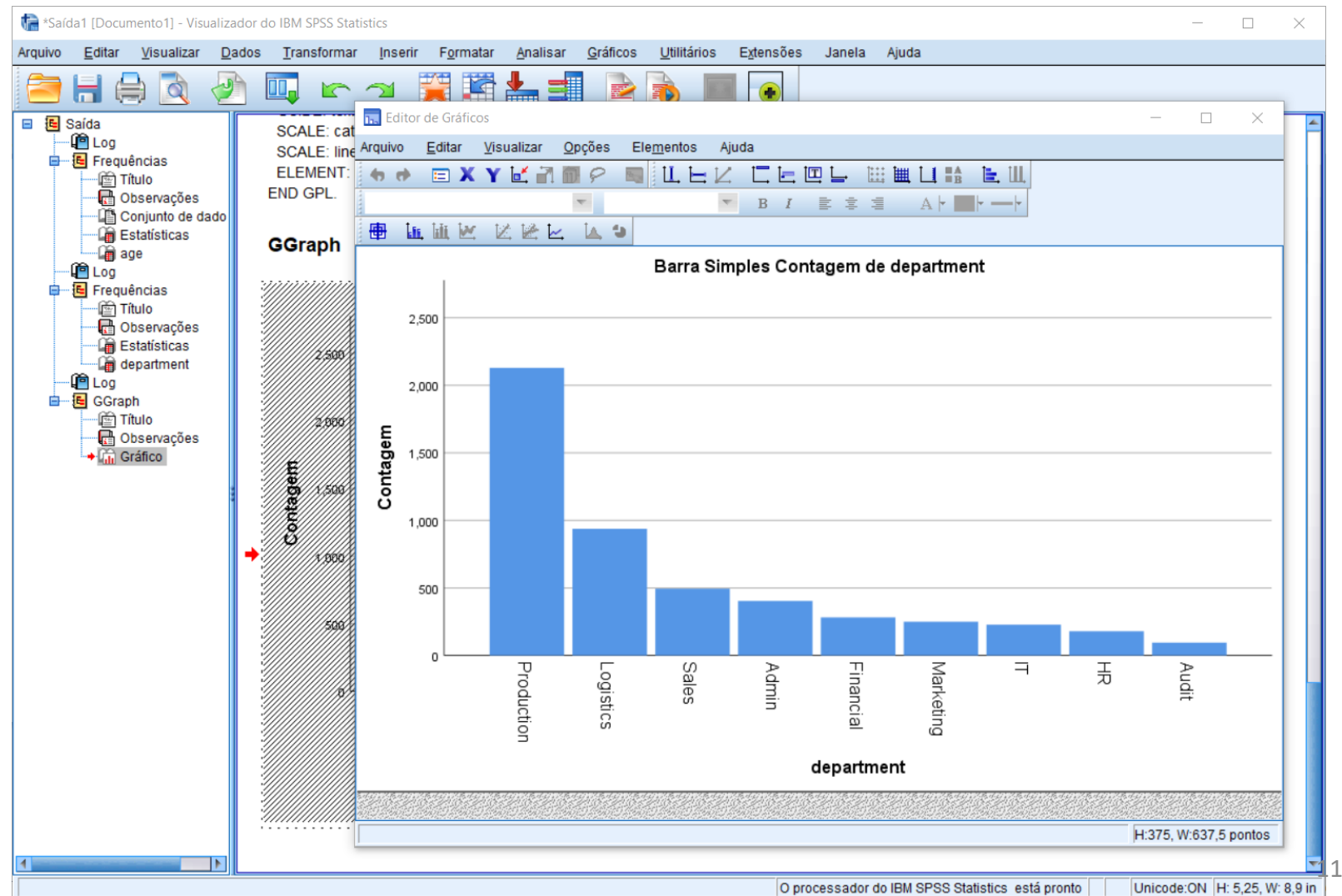


Gráfico de Barras

- O gráfico é publicado no 'Visualizador de Resultados'
- Para editarmos o gráfico, fazer duplo-clique em cima do mesmo
- Isso irá abrir o 'Editor de Gráficos'
- Entre outras coisas o 'Editor de Gráfico' permite mudar a cor das barras

