



Aula 5: As Mulheres Recebem o Mesmo que os Homens?

Percentagens, médias, medianas e outras medidas de estatística descritiva

Docente: Amílcar Moreira

Data & Hora: 22/10/2020, 20:30-22:30

Local: FRANCESINHAS 2, Sala 101

DESIGUALDADE

Mulheres ganham menos do que homens em todas as profissões na Europa

28/5/2018, 7:35 1.271 13

Segundo o estudo "Igualdade de Género ao Longo da Vida - Portugal no contexto europeu", as mulheres europeias são, em média, mais escolarizadas do que os homens, mas ganham salários mais baixos.

Partilhe     



Nova App NOS Net

Inteligente é uma net que faz o que queres

Gerir horários Mudar password Criar Wi-Fi visitas



A NOVA NET DA NOS É GIGA

Saber mais

ÚLTIMAS

14:38 Estoril -- Paços de Ferreira retoma segunda-feira

DESIGUALDADE

Mulheres ganham menos do que homens em todas as profissões na Europa

28/5/2018, 7:35 1.271 13

Segundo o estudo "Igualdade de Género ao Longo da Vida - Portugal no contexto europeu", as mulheres europeias são, em média, mais escolarizadas do que os homens, mas ganham salários mais baixos.

Partilhe     



Olhando para a disparidade salarial, Portugal apresenta um valor inferior à média da União Europeia a 27, já que as mulheres portuguesas **“ganham menos 11,2 euros/hora do que os homens”**, quando a média europeia é de 17,4 euros.

ÚLTIMAS

14:38 Estoril -- Paços de Ferreira retoma segunda-feira

- **Objetivos da Aula**

- **Parte Teórica**

- Saber identificar o que devemos procurar para conhecermos a variável que nos interessa
- Saber escolher qual o tipo de medidas mais adequadas em função da natureza da variável
- Perceber a importância dos dados extremos (*outliers*) na análise da distribuição das variáveis
- Saber escolher os tipos de gráficos mais adequados para visualizar a distribuição dos dados

- **Parte Prática**

- Produzir Tabelas de Frequências
- Produzir tabelas com medidas de tendência central e de dispersão
- Produzir gráficos para visualizar a distribuição das variáveis
- Limpar os dados: Identificar e codificar valores extremos



- **Como descrever a variável que nos interessa?**
 - **Começamos por olhar a medidas que sumarizem a distribuição da variável (*medidas de tendência central*)**
 - **Depois olhamos para a forma como os valores estão distribuídos na variável (*distribuição de frequências*)**
 - **Finalmente olhamos para o grau de variância na forma como se distribuem os valores da variável (*medidas de dispersão*)**



- **Como descrever a variável que nos interessa?**
- **Alguns conselhos práticos:**
 - **Começar sempre das medidas mais simples para as mais complexas/detalhadas**
 - **Escolher o tipo de medida em função da natureza da variável**
 - **Perceber se a existência de casos extremos (*outliers*) pode afetar a aplicação de medidas de Estatística Descritiva**

- **Medidas de tendência central**

	DEFINIÇÃO	TIPO DE VARIÁVEL	OBSERVAÇÕES
Moda (ou <i>valor modal</i>)	Valor mais frequente da variável	Todos	<ul style="list-style-type: none">• Adequada se a distribuição da variável é enviesada• Não é sensível a valores extremos
Média	Valor médio das observações na variável	Razão	<ul style="list-style-type: none">• 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)	Razão	<ul style="list-style-type: none">• Adequada se a distribuição da variável é enviesada• Não é sensível a valores extremos

- Distribuição das Frequências**

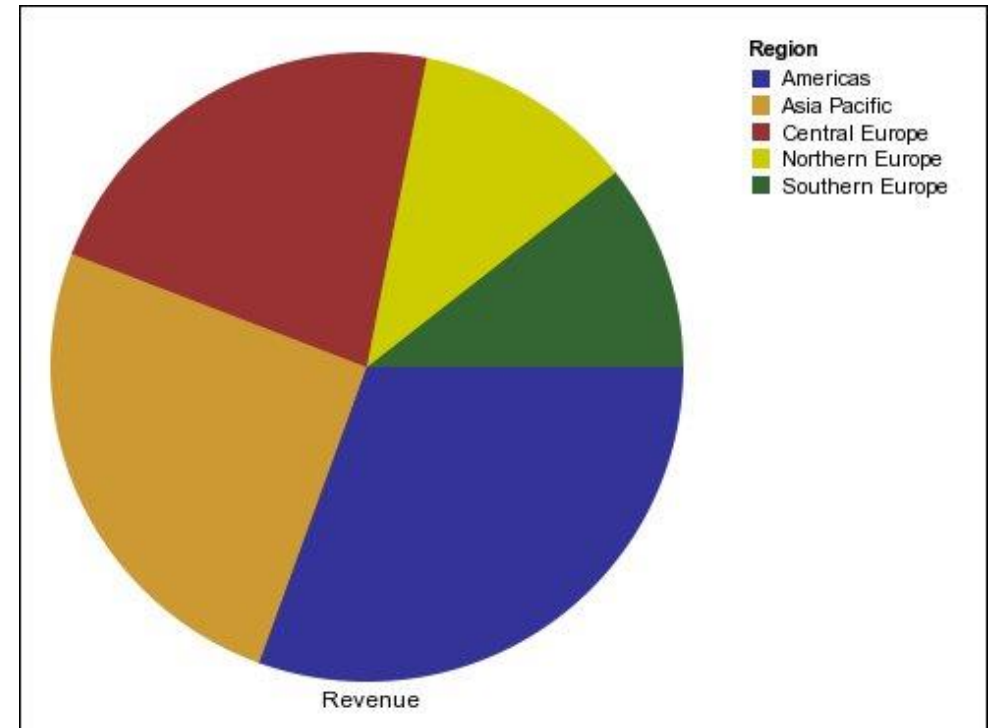
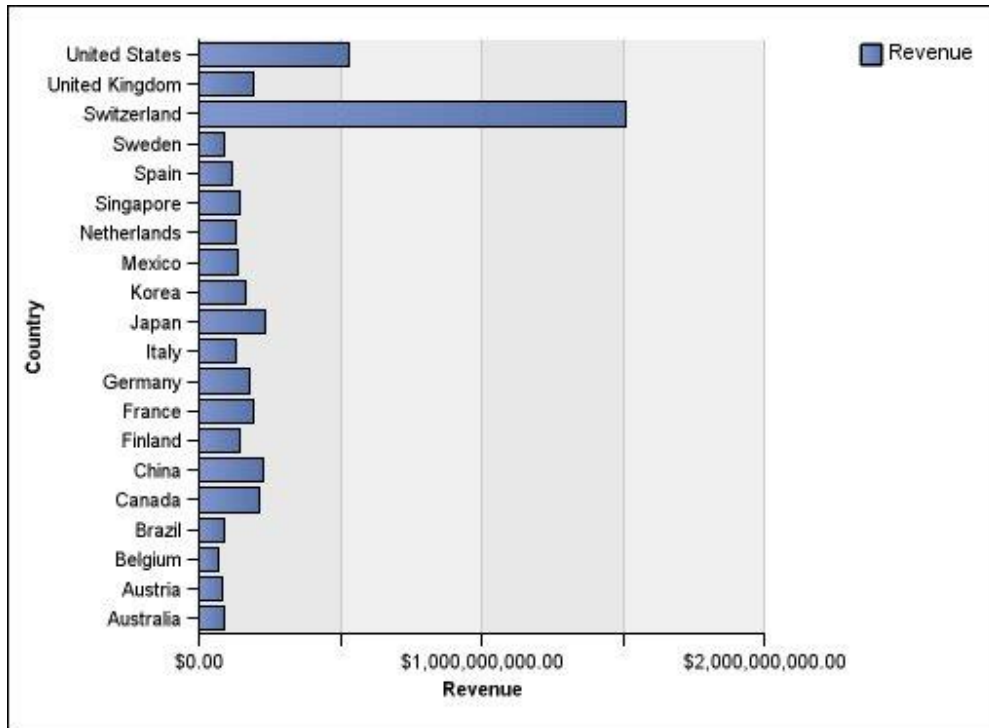
	DEFINIÇÃO	TIPO DE VARIÁVEL	OBSERVAÇÕES
Frequência	Numero observações com um determinado valor na variável	Todos	-
Proporção	Número observações com um determinado valor na variável, dividido pelo número total de observações.	Todos	• Expressa numa escala de 0-1
Percentagem	Número observações com um determinado 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)

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

Variável
Nominal
ou Ordinal

Gráfico de Barras

Gráfico Circular

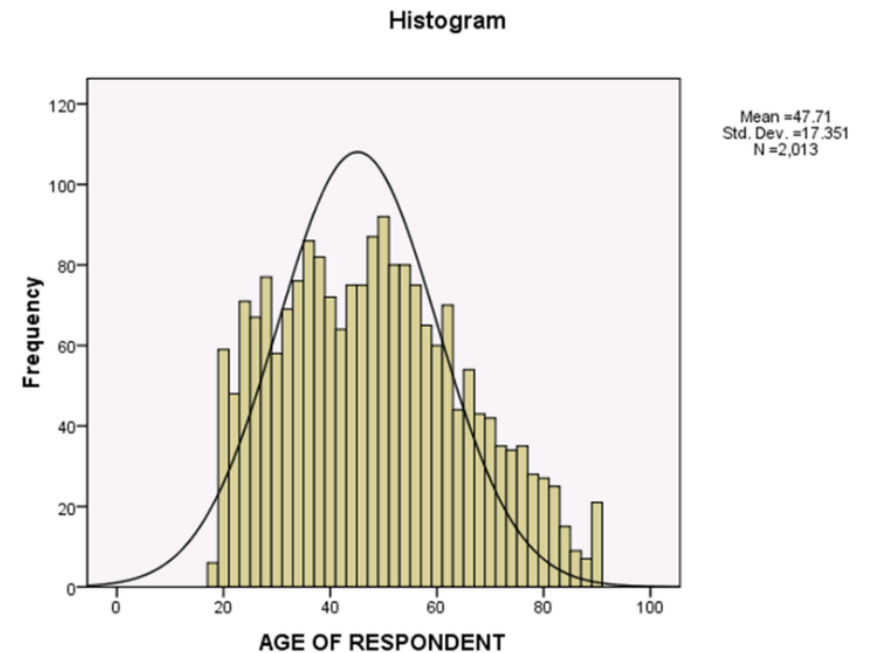
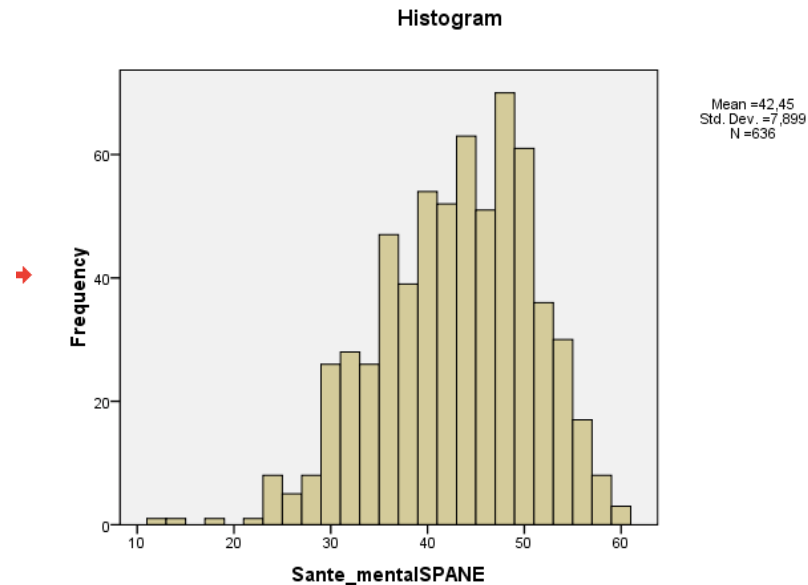


- Qual a melhor forma de visualizar a distribuição de frequências?

Variável
Intervalar
ou Razão

Histograma

Histograma,
com Curva de Distribuição



- **Medidas de Dispersão**

- **Três tipos**

- Baseadas em valores absolutos (ex: Intervalo de Variação)
- Calculadas em função de percentis/ decis / quartis (ex: Amplitude Interquartis, Rácios interdecis 90/10 ou 80/20)
- Calculadas em função da média (Variância, Desvio Padrão)

- **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

- **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
 - **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
 - **Quartil – Apura-se dividindo a variável em 10 partes iguais.**
 - 1º Quartil - ex: 25% das pessoas têm salários até €600
 - 3º Quartil - ex: 75% das pessoas têm salários até €1500

- **Medidas de dispersão**

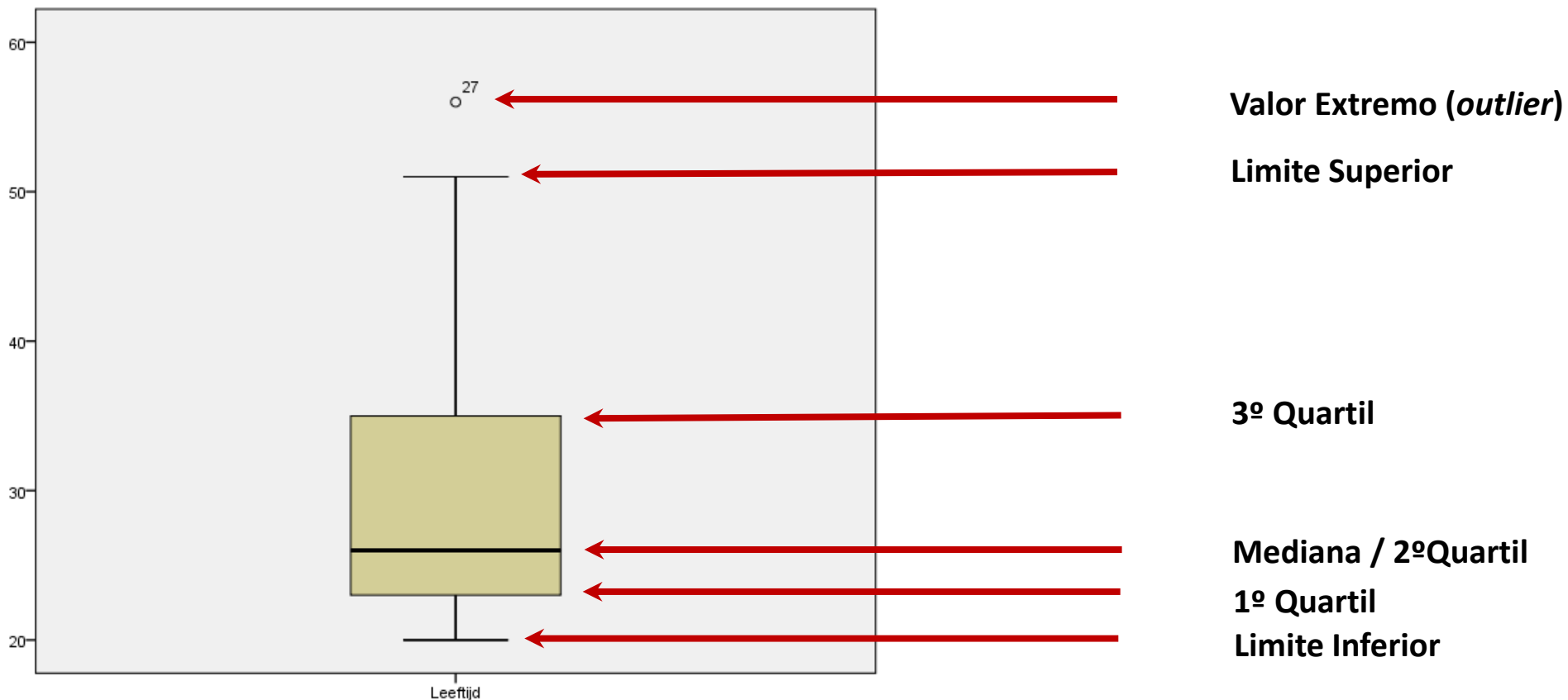
	DEFINIÇÃO	TIPO DE VARIÁVEL	OBSERVAÇÕES
Intervalo de Variação	Diferença entre o valor mais alto e o valor mais baixo da variável	Intervalar* Razão	<ul style="list-style-type: none"> • É sensível a valores extremos
Amplitude Interquartis	Diferença entre o valor do 3º Quartil e o valor do 1º Quartil	Intervalar* Razão	<ul style="list-style-type: none"> • Não é sensível a valores extremos
Rácio Interdecis	90/10: Valor do 9º decil dividido pelo valor do 1º decil	Intervalar* Razão	<ul style="list-style-type: none"> • Alternativa: 80/20 • Não é sensível a valores extremos

- **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	<ul style="list-style-type: none"> • É sensível a valores extremos
Amplitude Interquartis	Diferença entre o valor do 3º Quartil e o valor do 1º Quartil	Intervalar* Razão	<ul style="list-style-type: none"> • Não é sensível a valores extremos
Rácio Interdecis	90/10: Valor do 9º decil dividido pelo valor do 1º decil	Intervalar* Razão	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • = 0: Não há variância • Quanto maior, maior variância • É sensível a valores extremos • Difícil de interpretar
Desvio-Padrão	Raiz quadrada da variância.	Intervalar* Razão	<ul style="list-style-type: none"> • Mais fácil de interpretar • Usa a mesma escala de valores da variável • É sensível a valores extremos

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

Gráfico de Caixa



Análise Exploratória dos Dados

Medidas de Tendências Central

Distribuição das Frequências

Medidas de Dispersão

Análise Exploratória dos Dados

Medidas de Tendências Central

Medidas de Tendência Central

- **Objectivo:**
 - **Produzir um conjunto de indicadores que sumarizam a distribuição etária dos trabalhadores na organização**

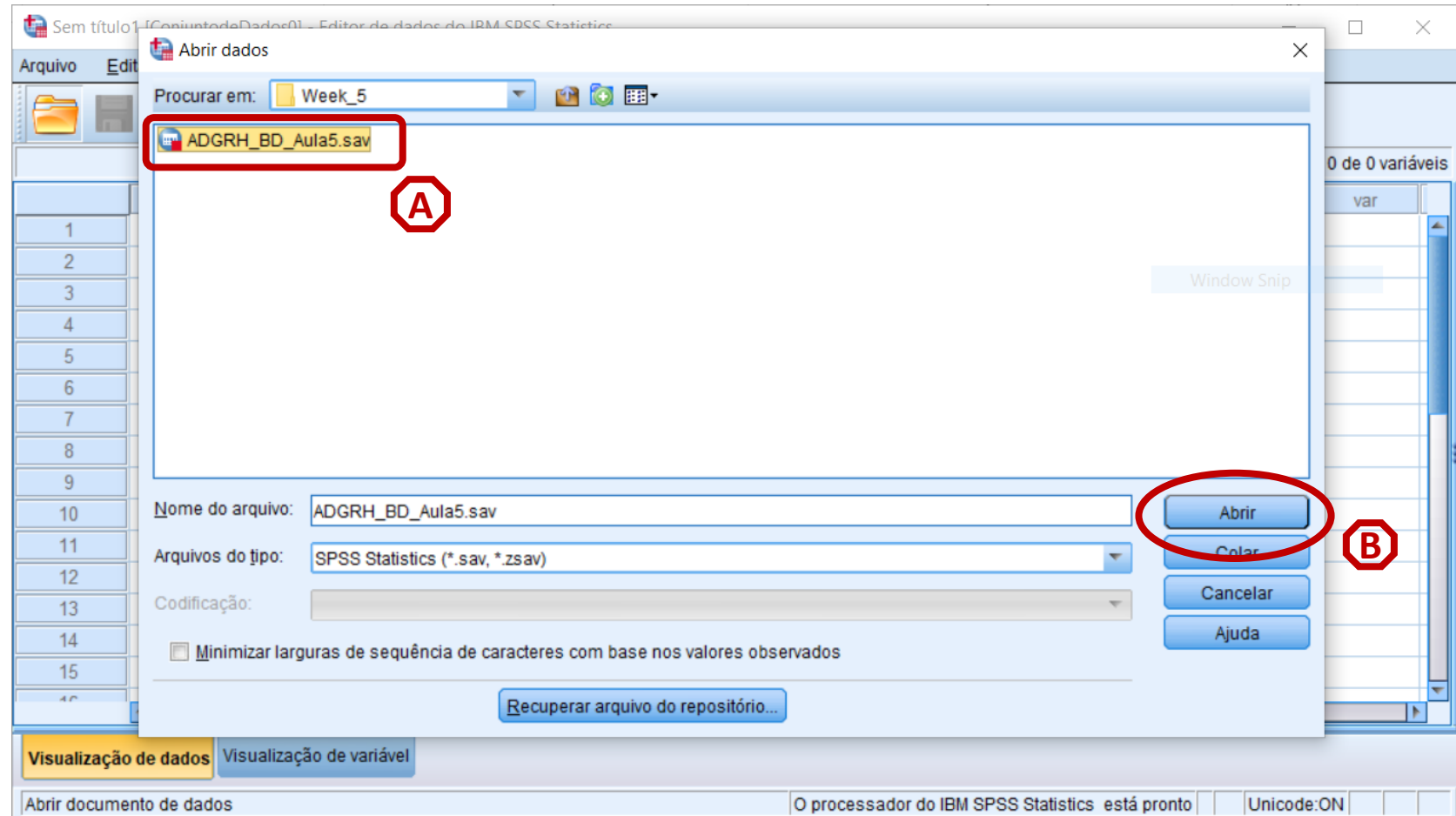
Média

Mediana

Moda

Medidas de Tendência Central

- Seleccionar ficheiro 'ADGRH_BD_Aula5.sav'
- Seleccionar 'Abrir'



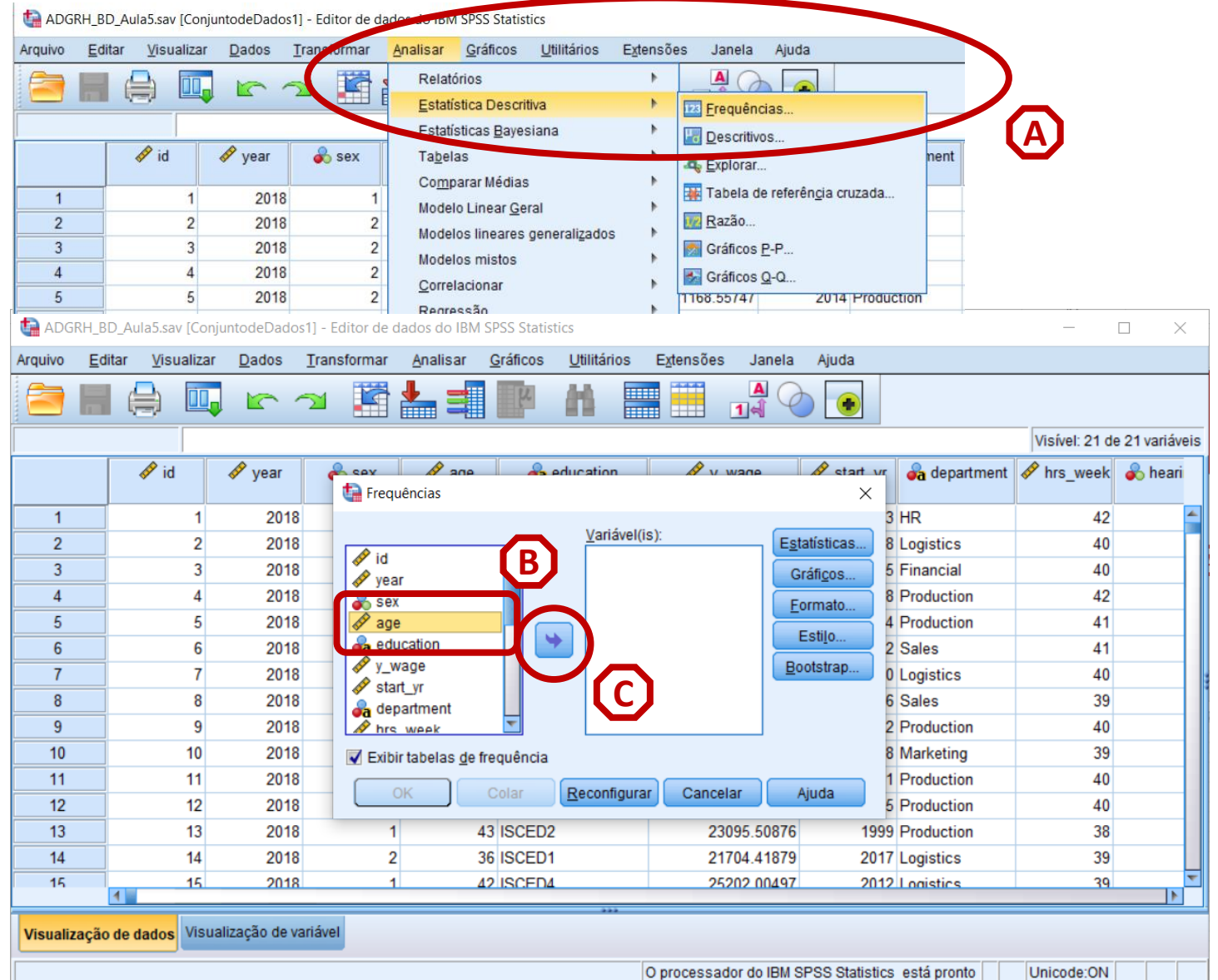
Medidas de Tendência Central

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

A

B

C



ADGRH_BD_Aula5.sav [Conjunto de Dados] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar **Analisar** Gráficos Utilitários Extensões Janela Ajuda

Relatórios
 Estatística Descritiva
 Estatísticas Bayesianas
 Tabelas
 Comparar Médias
 Modelo Linear Geral
 Modelos lineares generalizados
 Modelos mistos
 Correlacionar
 Regressão

1168.55747 2014 Production

ADGRH_BD_Aula5.sav [Conjunto de Dados] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar **Analisar** Gráficos Utilitários Extensões Janela Ajuda

id year sex age education y_wage start_yr department hrs_week

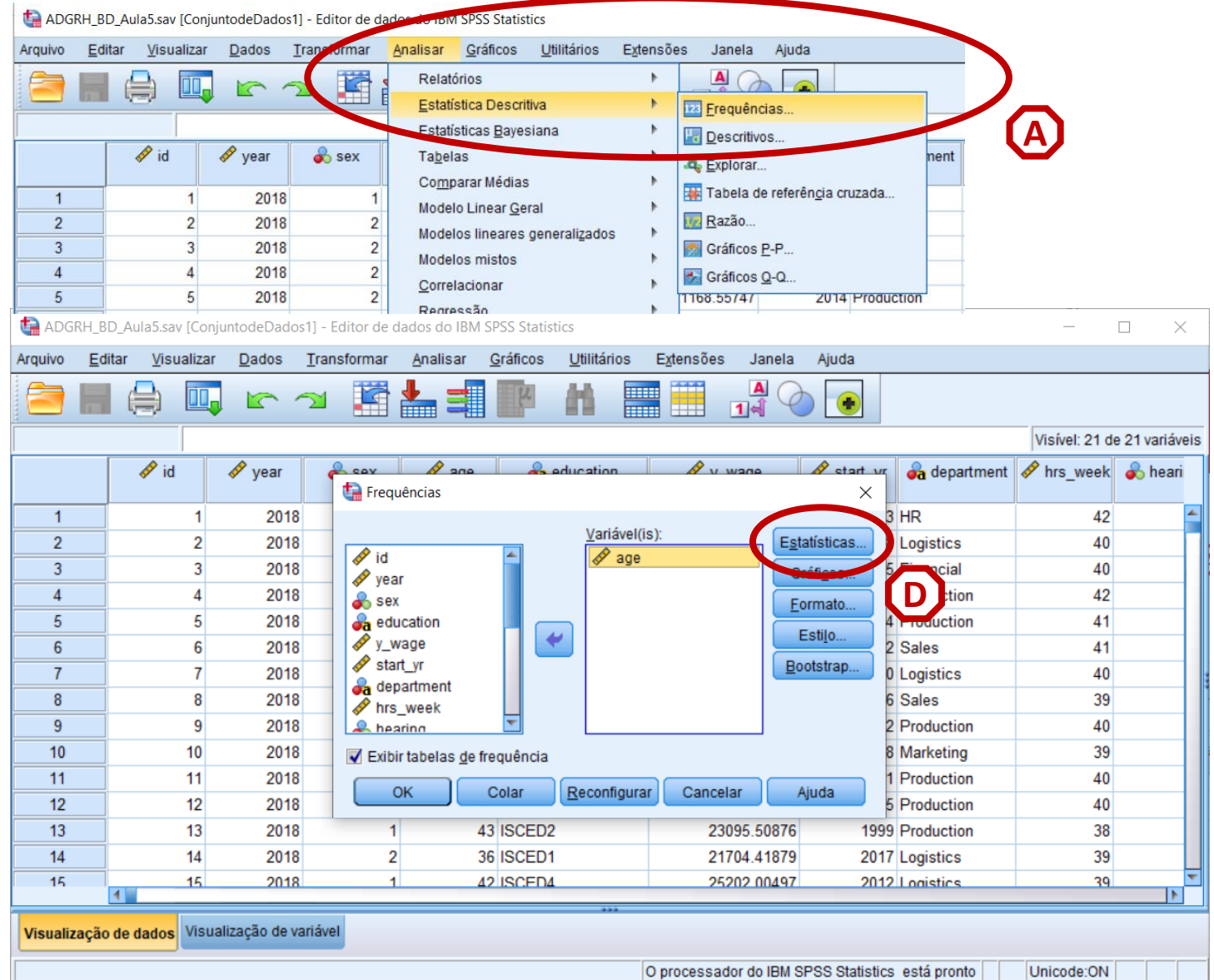
1 1 2018 1
 2 2 2018 2
 3 3 2018 2
 4 4 2018 2
 5 5 2018 2
 6 6 2018 2
 7 7 2018 2
 8 8 2018 2
 9 9 2018 2
 10 10 2018 2
 11 11 2018 2
 12 12 2018 2
 13 13 2018 1 43 ISCED2 23095.50876 1999 Production 38
 14 14 2018 2 36 ISCED1 21704.41879 2017 Logistics 39
 15 15 2018 1 42 ISCED4 25202.00497 2012 Logistics 39

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

Medidas de Tendência Central

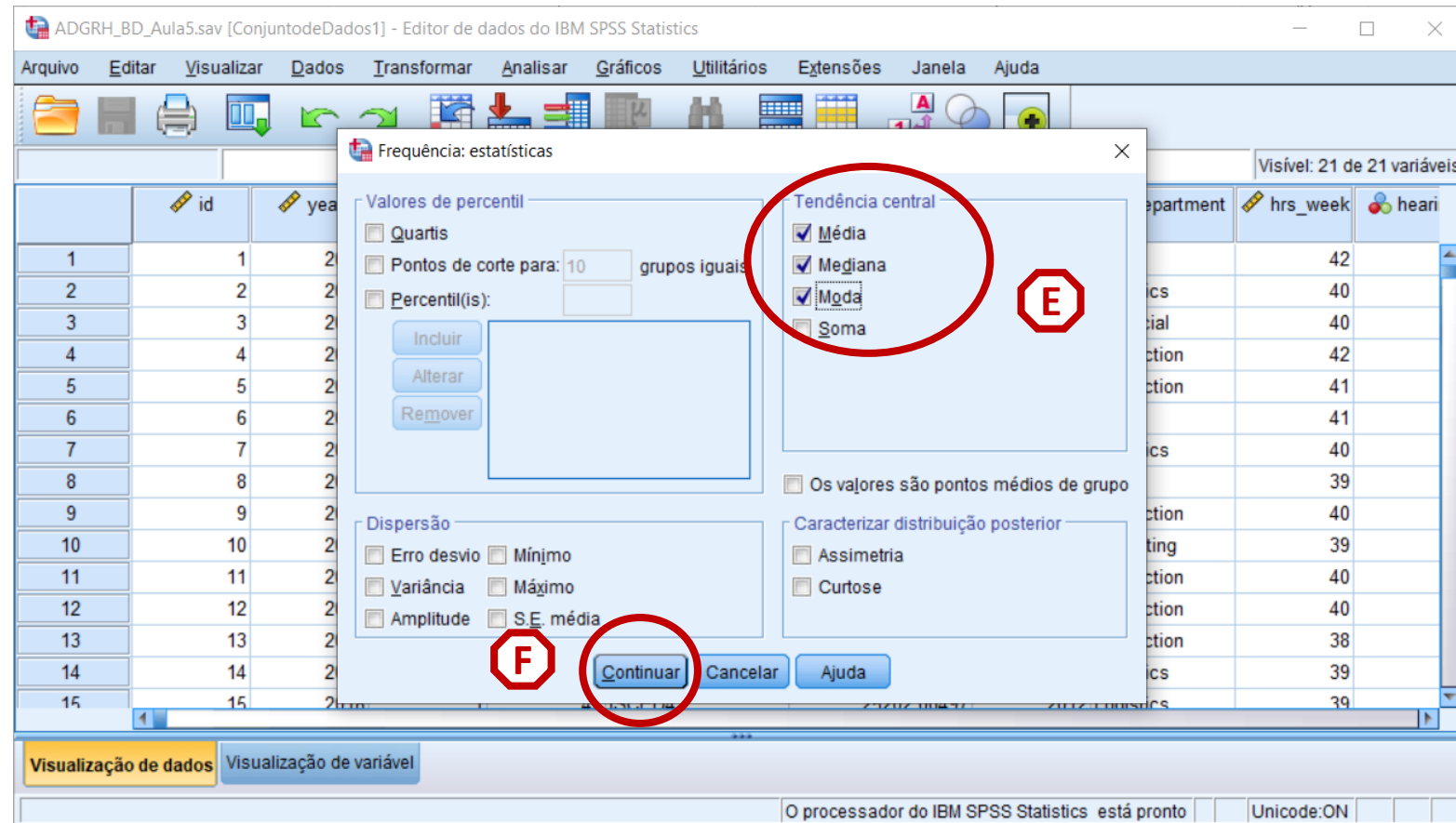
- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Frequências' A
- Selecionar variável 'age' B
- Colocar na caixa de 'Variáveis' C
- Selecionar 'Estatísticas' D



The image displays two screenshots of the IBM SPSS Statistics interface. The top screenshot shows the 'Analisar' (Analyze) menu with 'Estatística Descritiva' (Descriptive Statistics) > 'Frequências...' (Frequencies...) selected, highlighted with a red circle and a red 'A' in a hexagon. The bottom screenshot shows the 'Frequências' dialog box with 'age' selected in the 'Variável(is):' (Variable(s)) field and 'Estatísticas...' (Statistics...) selected in the options list, also highlighted with a red circle and a red 'D' in a hexagon. The background shows a data table with columns for id, year, sex, department, and hrs_week.

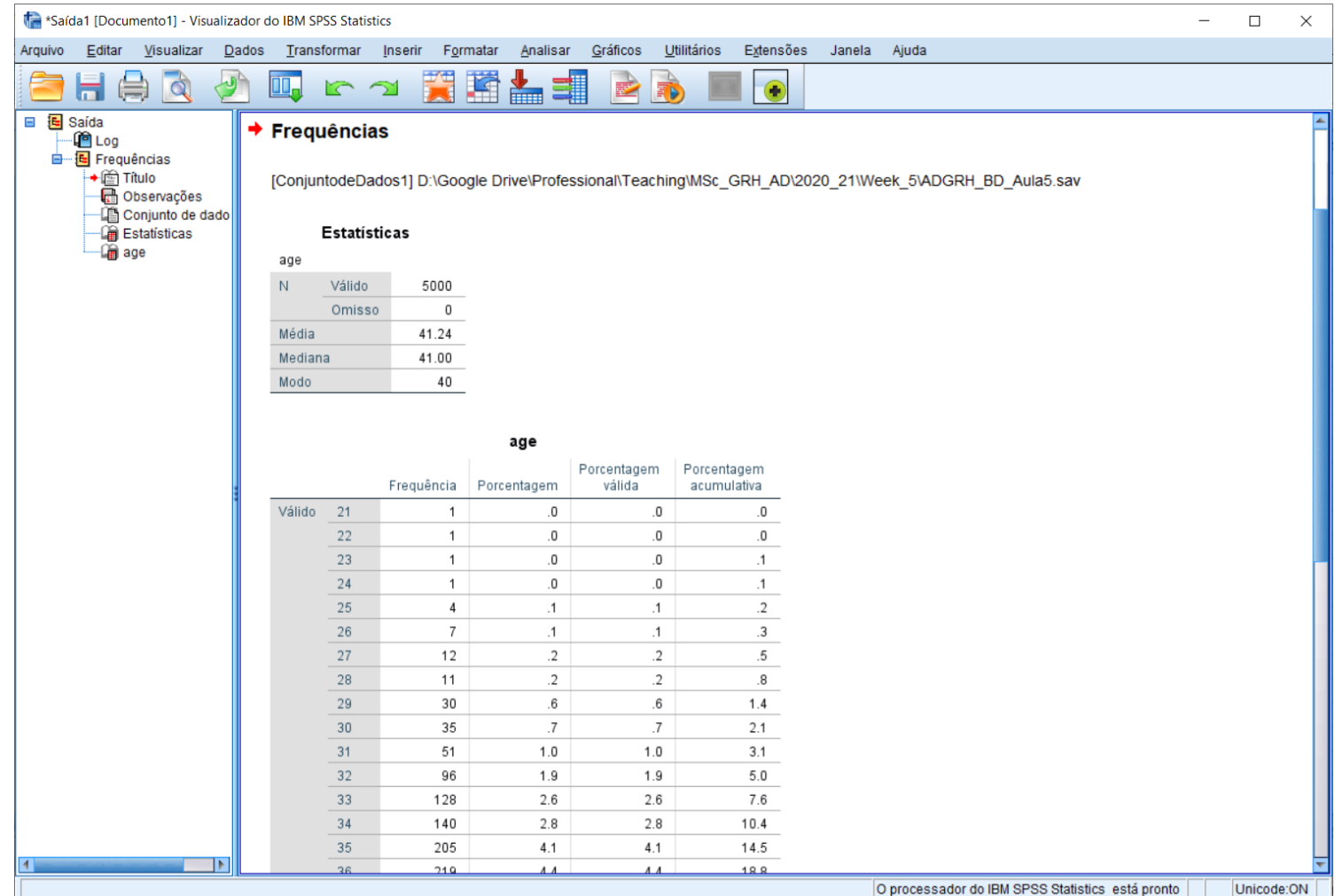
Medidas de Tendência Central

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Frequências' **A**
- Selecionar variável 'age' **B**
- Colocar na caixa de 'Variáveis' **C**
- Selecionar 'Estatísticas' **D**
- Selecionar Medidas de Tendência Central **E**
- Selecionar 'Continuar' / 'OK' **F**



Medidas de Tendência Central

- O resultado é publicado no 'Visualizador de Resultados'
- Média = 41.24
- Mediana = 41
- Moda = 40



The screenshot shows the IBM SPSS Statistics 'Visualizador de Resultados' window. The main panel displays the 'Frequências' (Frequencies) output for the variable 'age'. It includes a summary table of statistics and a detailed frequency table.

Estadísticas

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

age

	Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido	21	.0	.0	.0
	22	.0	.0	.0
	23	.0	.0	.1
	24	.0	.0	.1
	25	.1	.1	.2
	26	.1	.1	.3
	27	.2	.2	.5
	28	.2	.2	.8
	29	.6	.6	1.4
	30	.7	.7	2.1
	31	1.0	1.0	3.1
	32	1.9	1.9	5.0
	33	2.6	2.6	7.6
	34	2.8	2.8	10.4
	35	4.1	4.1	14.5
	36	4.4	4.4	18.8

Análise Exploratória dos Dados

Distribuição das Frequências

Distribuição de Frequências

- **Objectivo:**
 - **Perceber como se distribuem os trabalhadores pelos departamentos da empresa?**

Número de trabalhadores, por departamento (Frequências Absolutas)

Percentagem de trabalhadores, por departamento (Frequências Relativas)

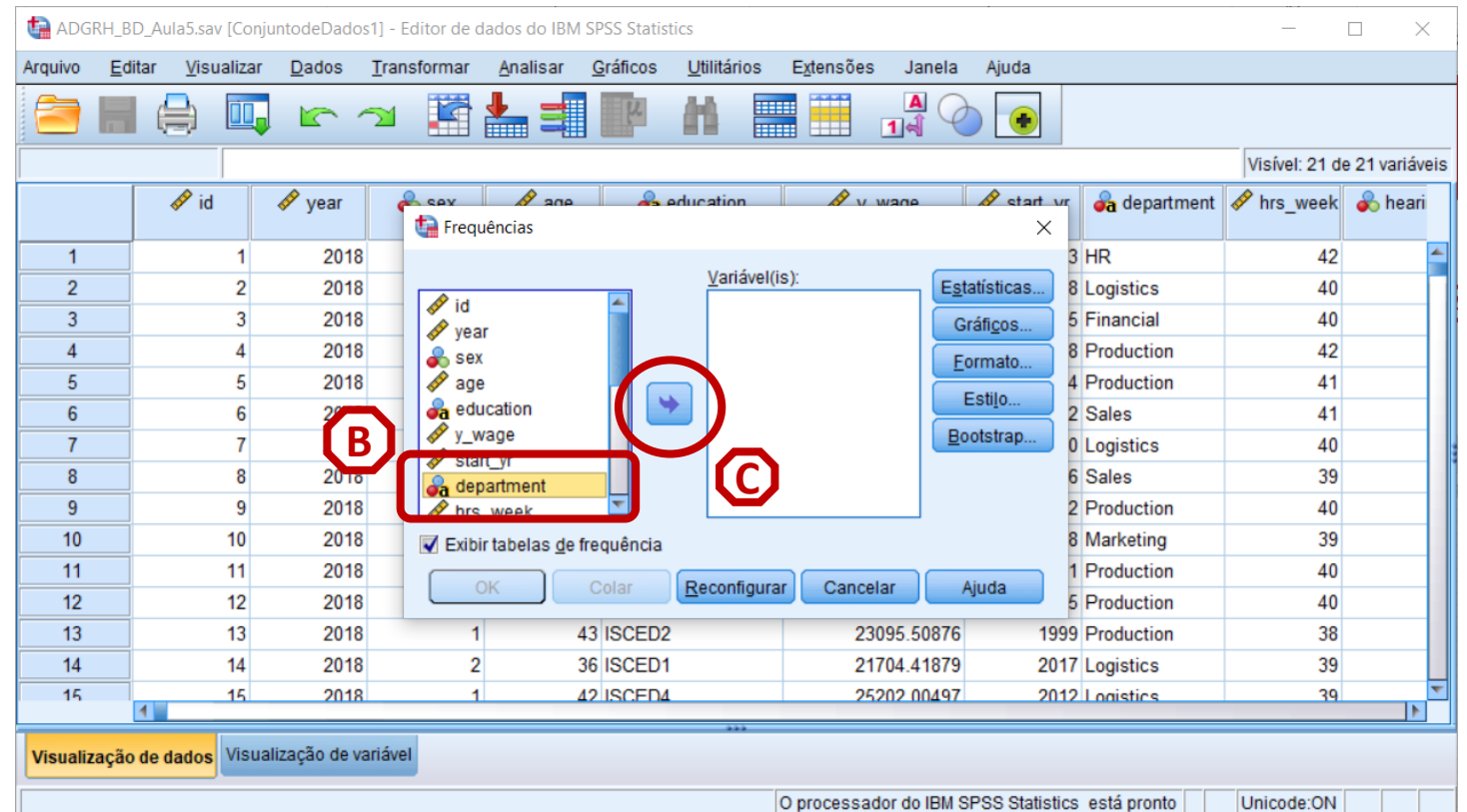
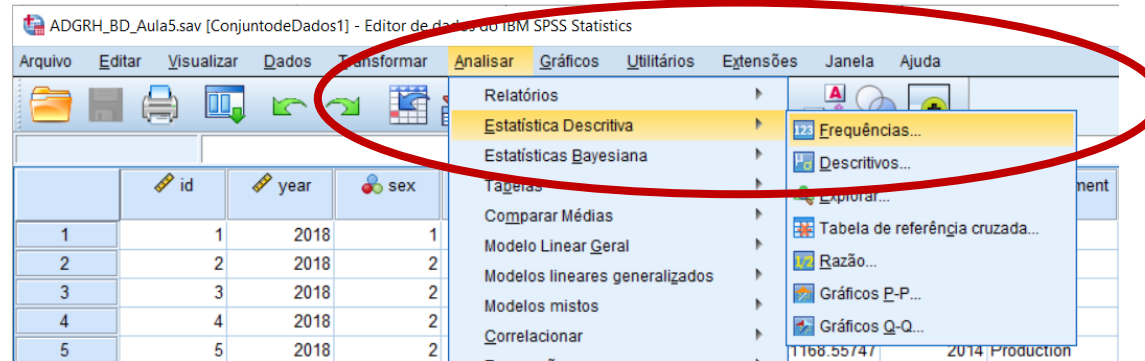
Distribuição de Frequências

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

A

B

C



Distribuição de Frequências

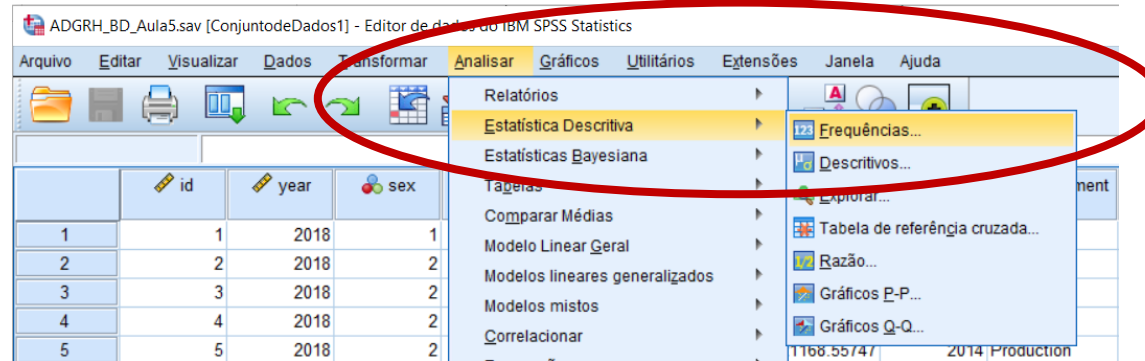
- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Frequências'
- Selecionar variável 'department'
- Colocar na caixa de 'Variáveis'
- Selecionar 'OK'

A

B

C

D



A

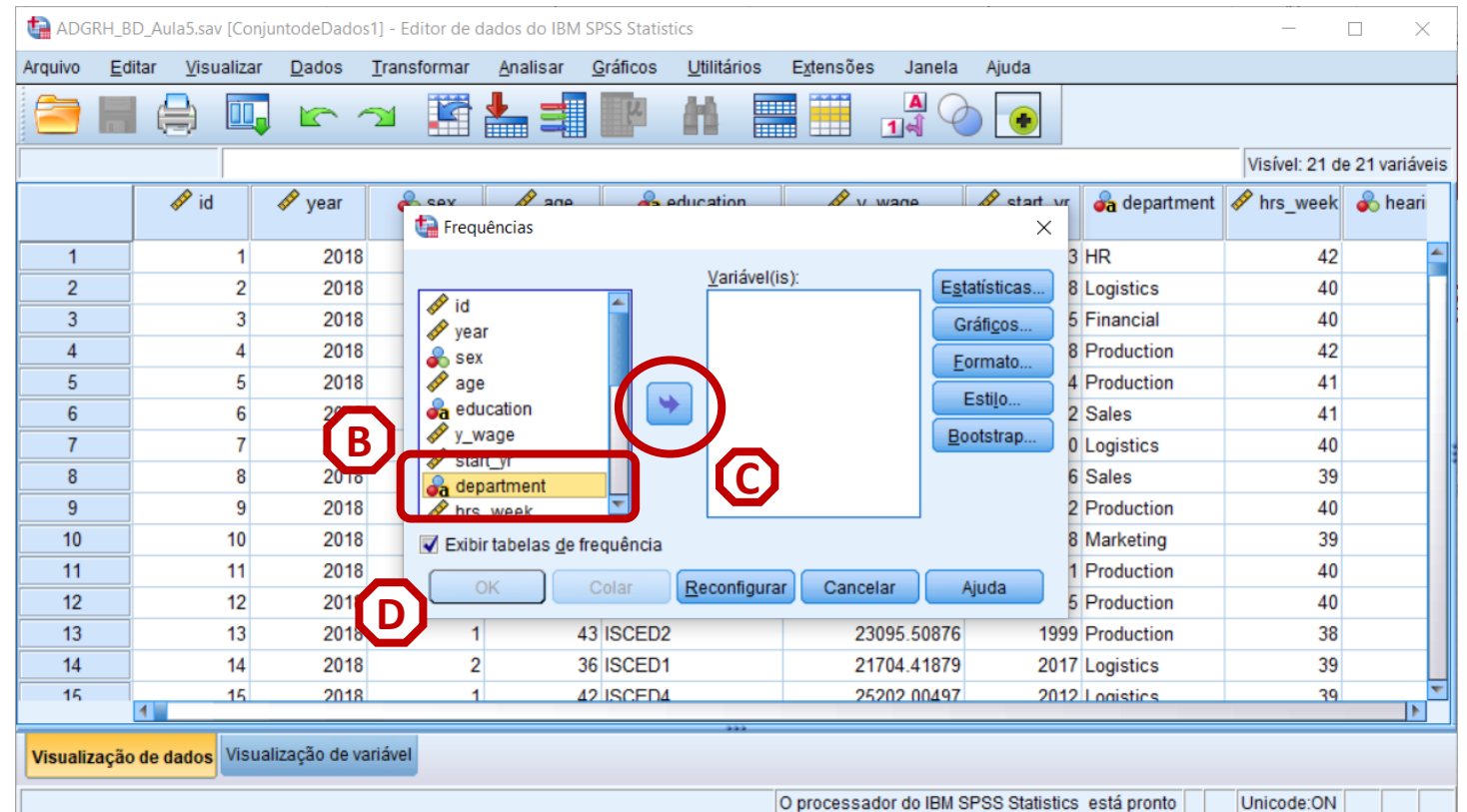
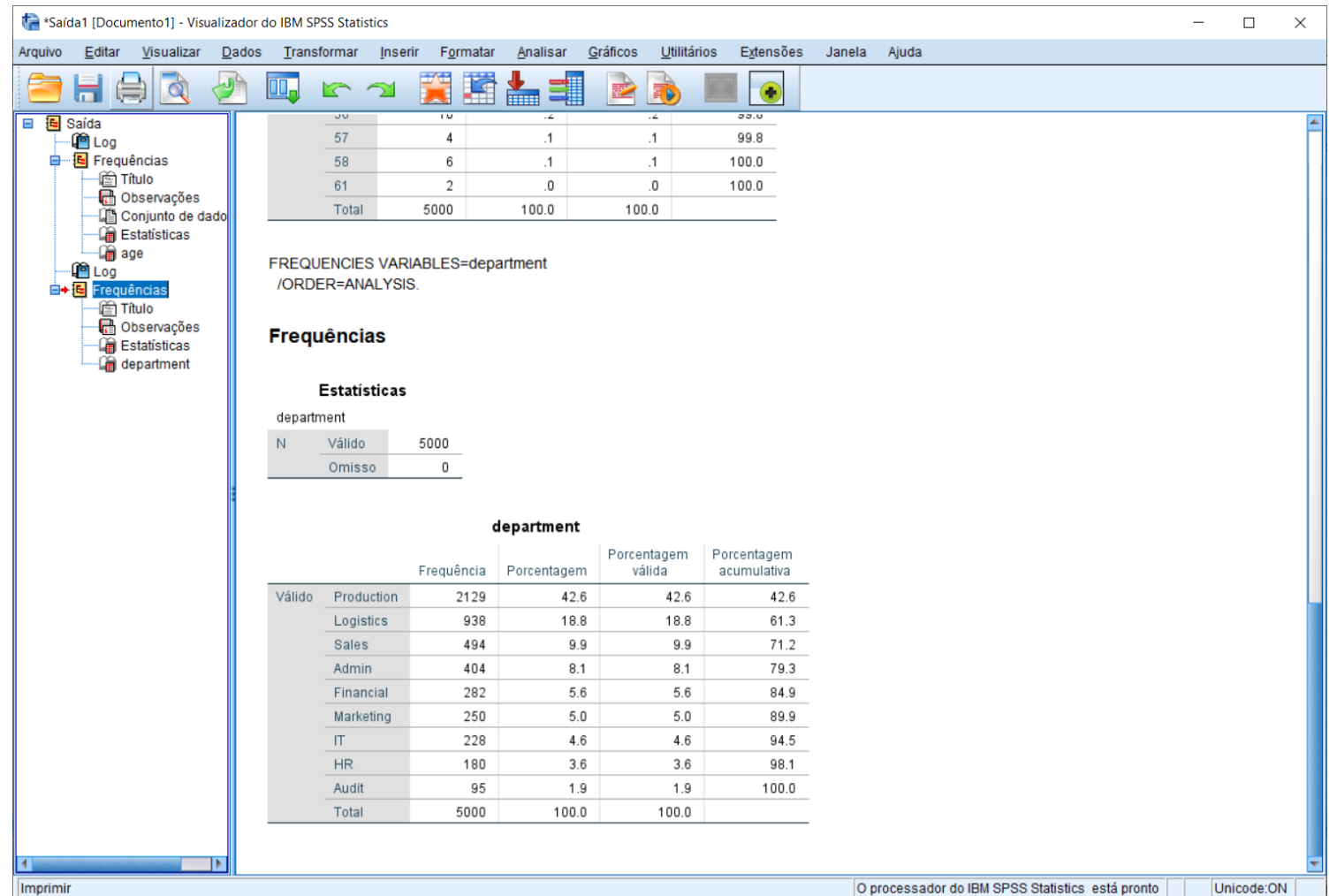


Tabela de Frequências

- O resultado é publicado no 'Visualizador de Resultados'



*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics
 Arquivo Editar Visualizar Dados Transformar Inserir Formatar Analisar Gráficos Utilitários Extensões Janela Ajuda

57	4	.1	.1	99.8
58	6	.1	.1	100.0
61	2	.0	.0	100.0
Total	5000	100.0	100.0	

FREQUENCIES VARIABLES=department
 /ORDER=ANALYSIS.

Frequências

Estatísticas

department

N	Válido	5000
	Omisso	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

Distribuição de Frequências

- **QUAL É A MELHOR FORMA DE VISUALIZAR ESTA DISTRIBUIÇÃO DE FREQUÊNCIAS?**

GRÁFICO DE BARRAS

GRÁFICO CIRCULAR (QUEIJO)

Gráfico de Barras

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

A

B

C

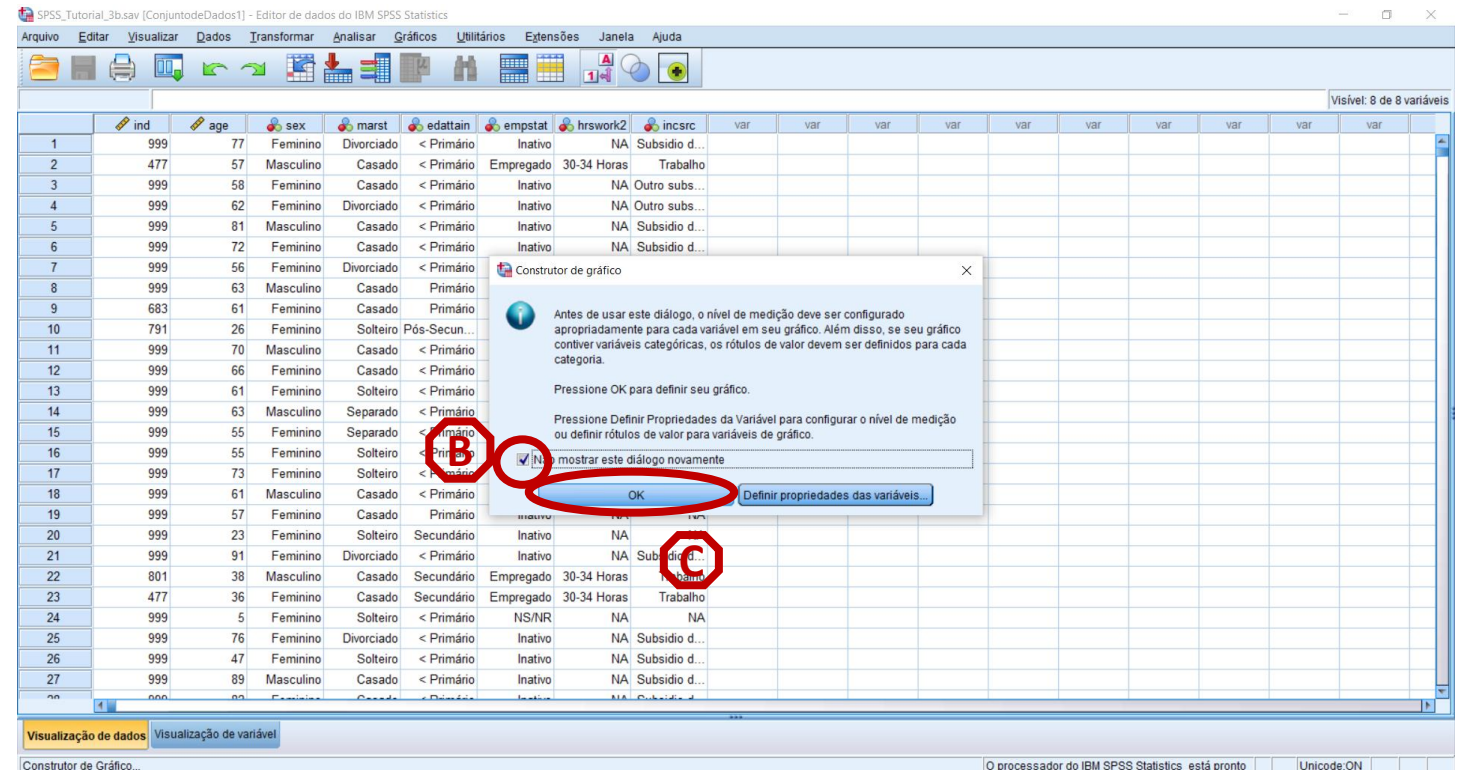
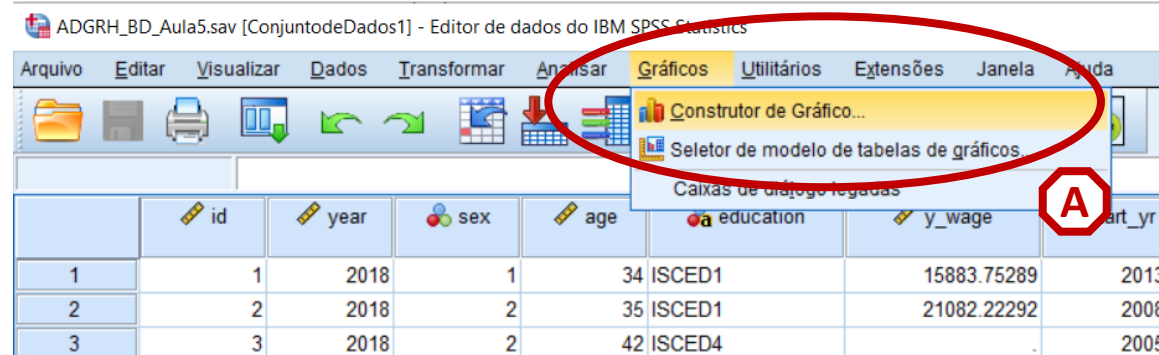


Gráfico de Barras

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

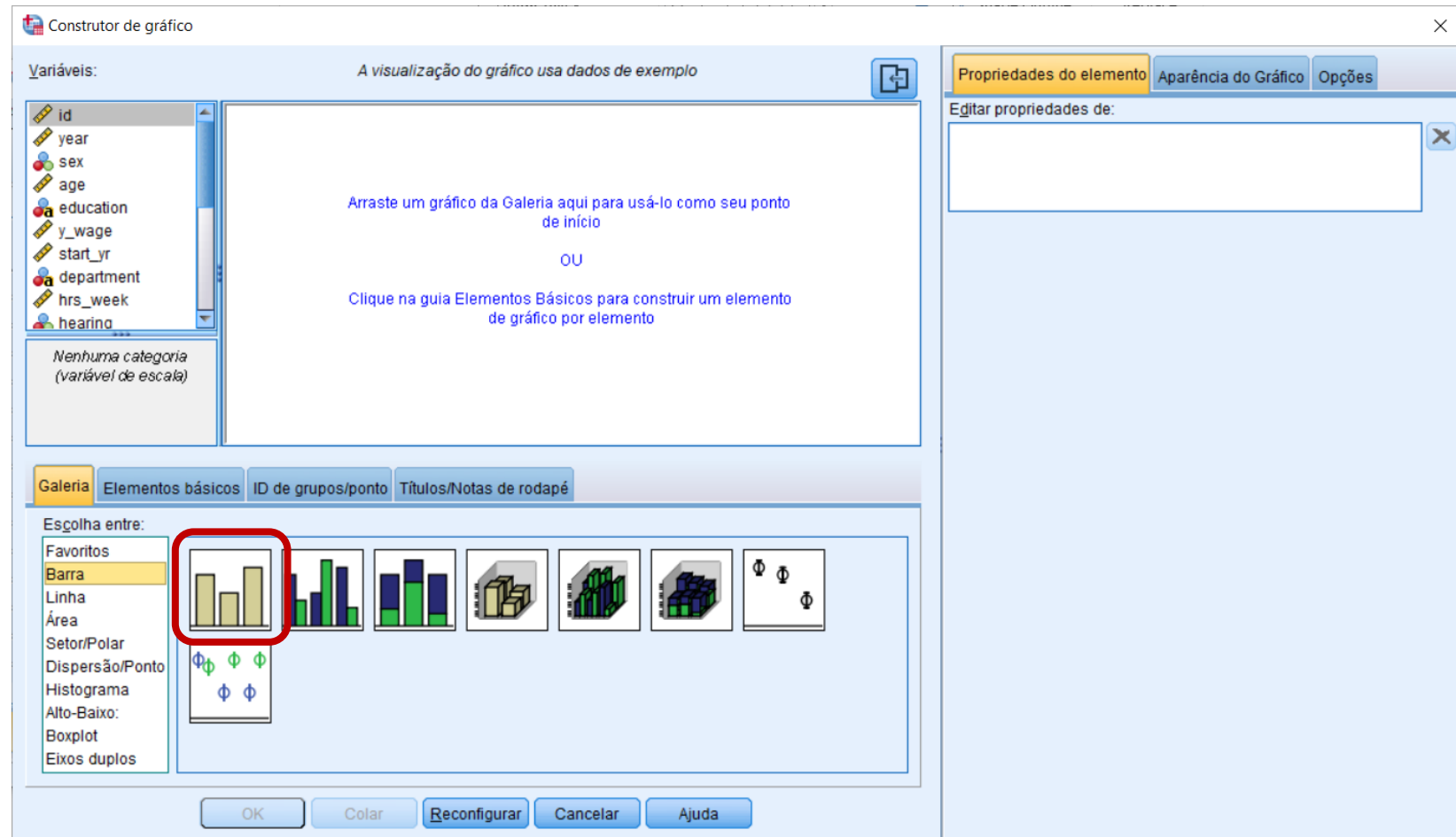


Gráfico de Barras

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

D

E

F

G

Gráfico de Barras

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

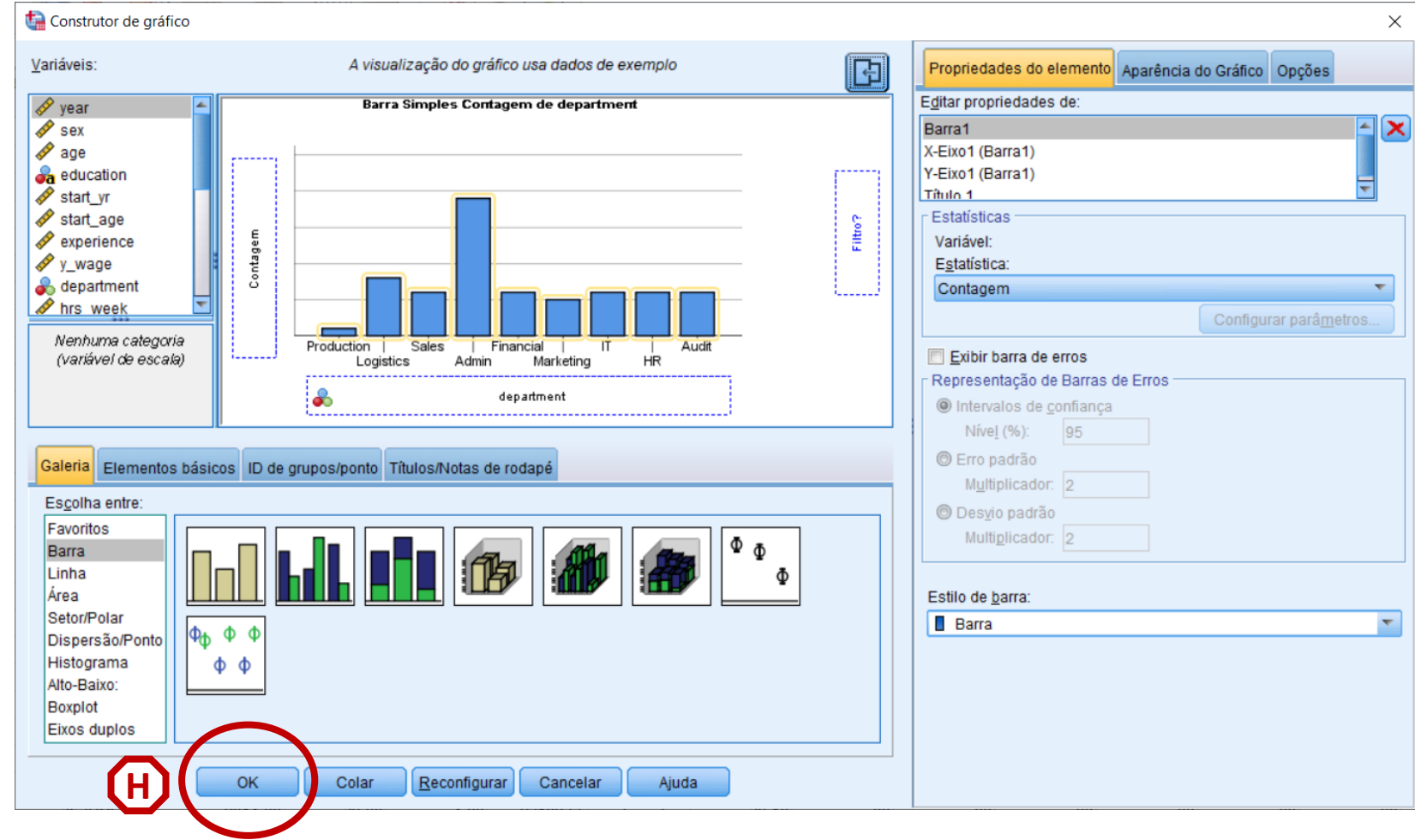


Gráfico de Barras

- O gráfico é publicado no 'Visualizador de Resultados'

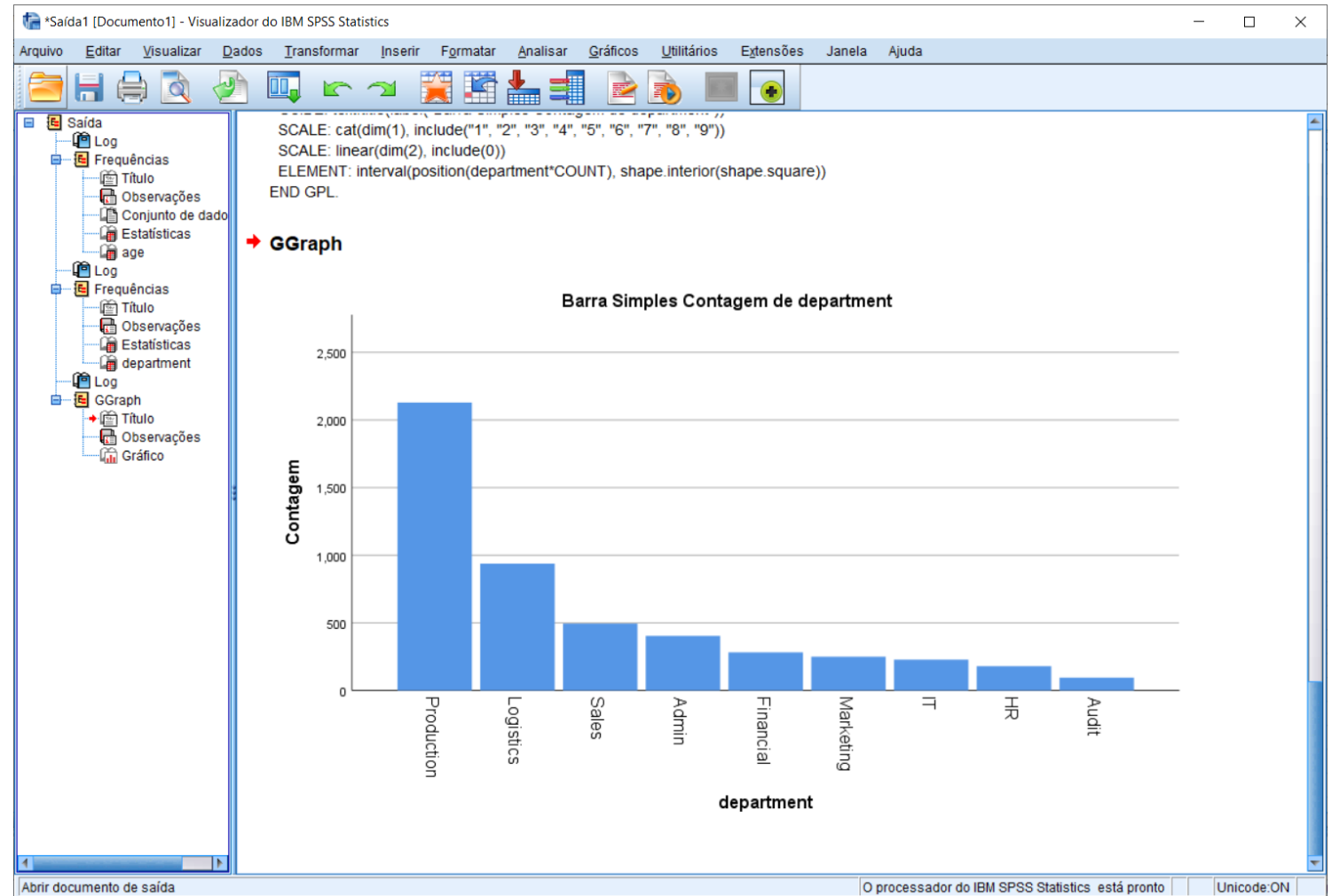


Gráfico de Barras

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

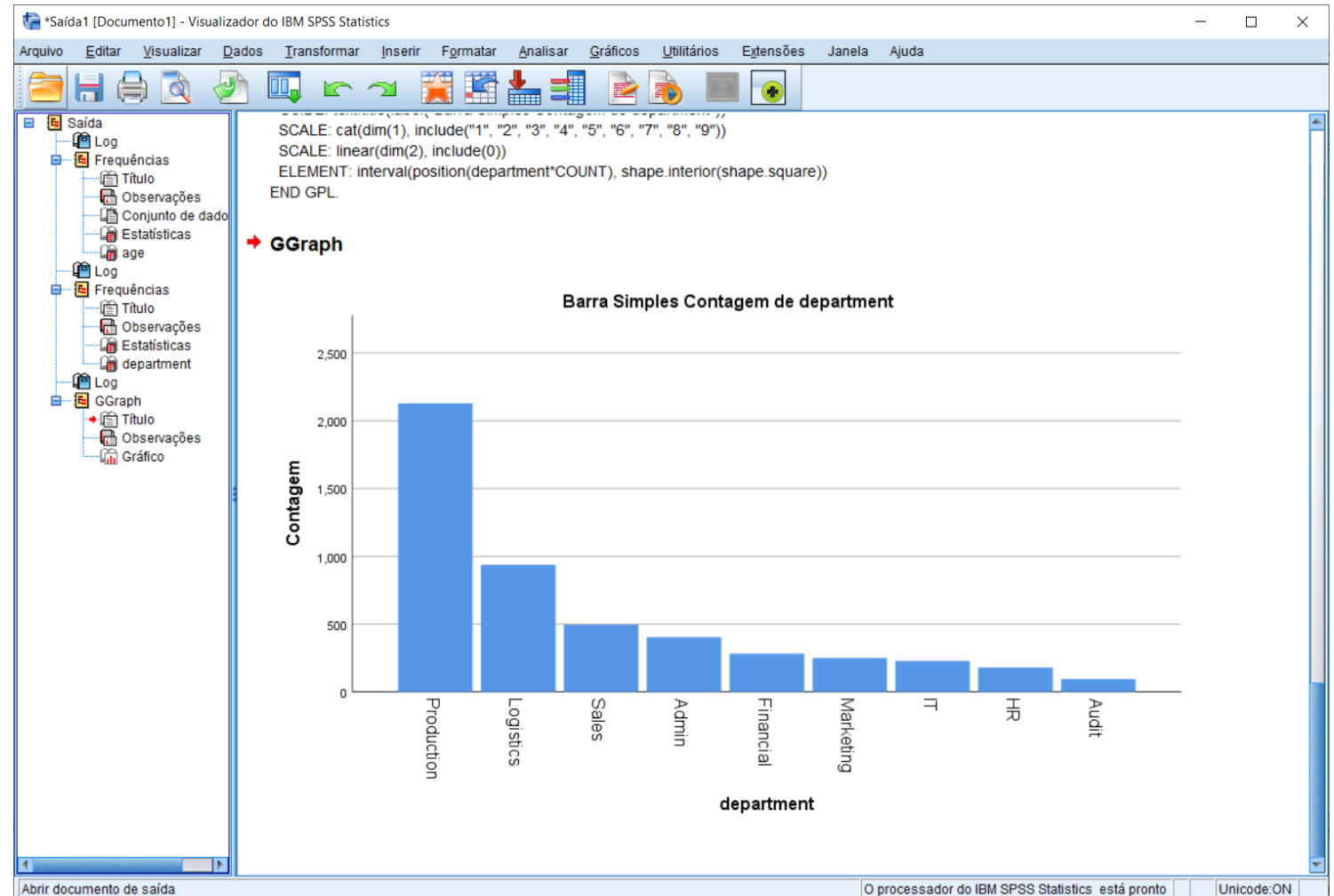


Gráfico de Barras

- 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'

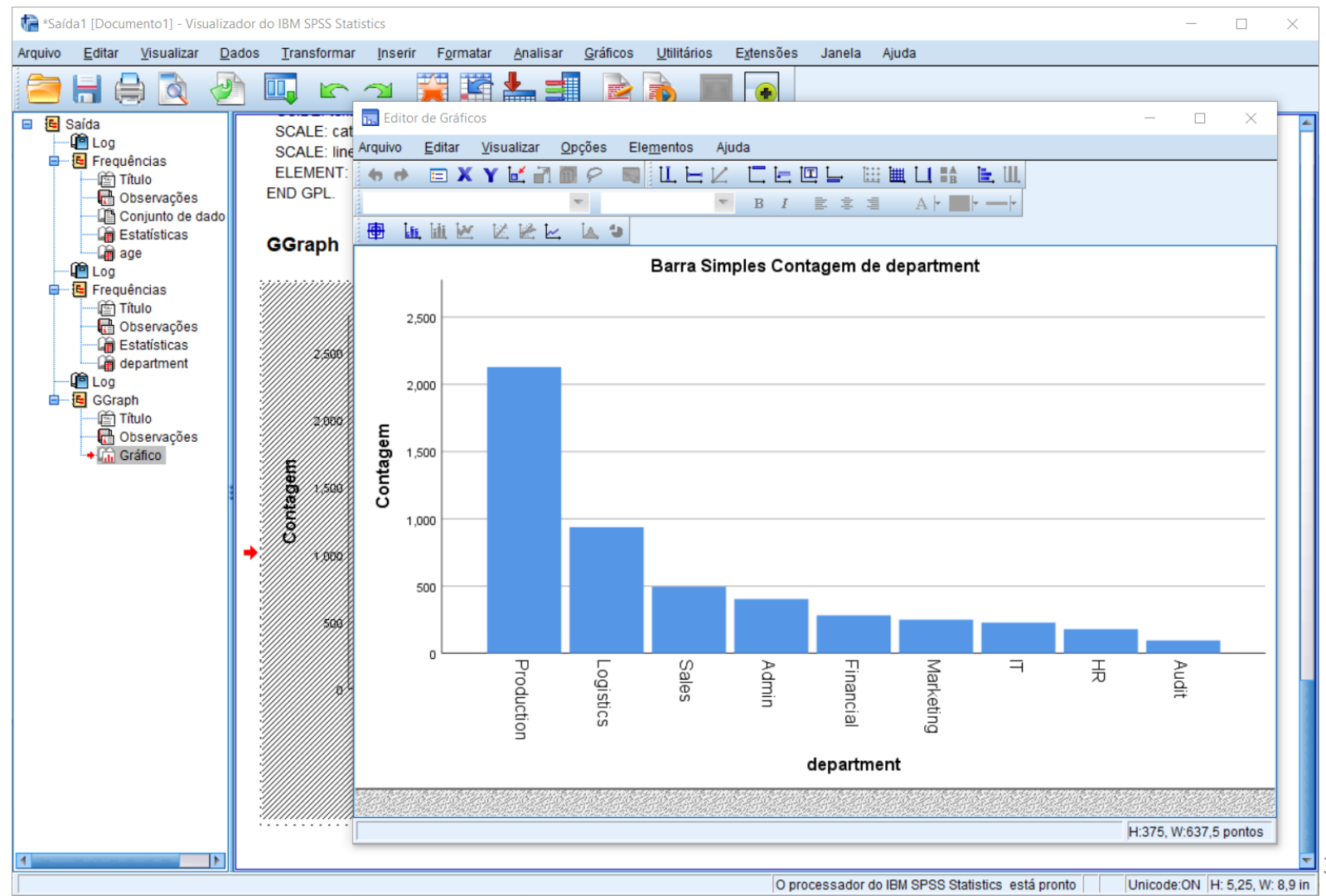
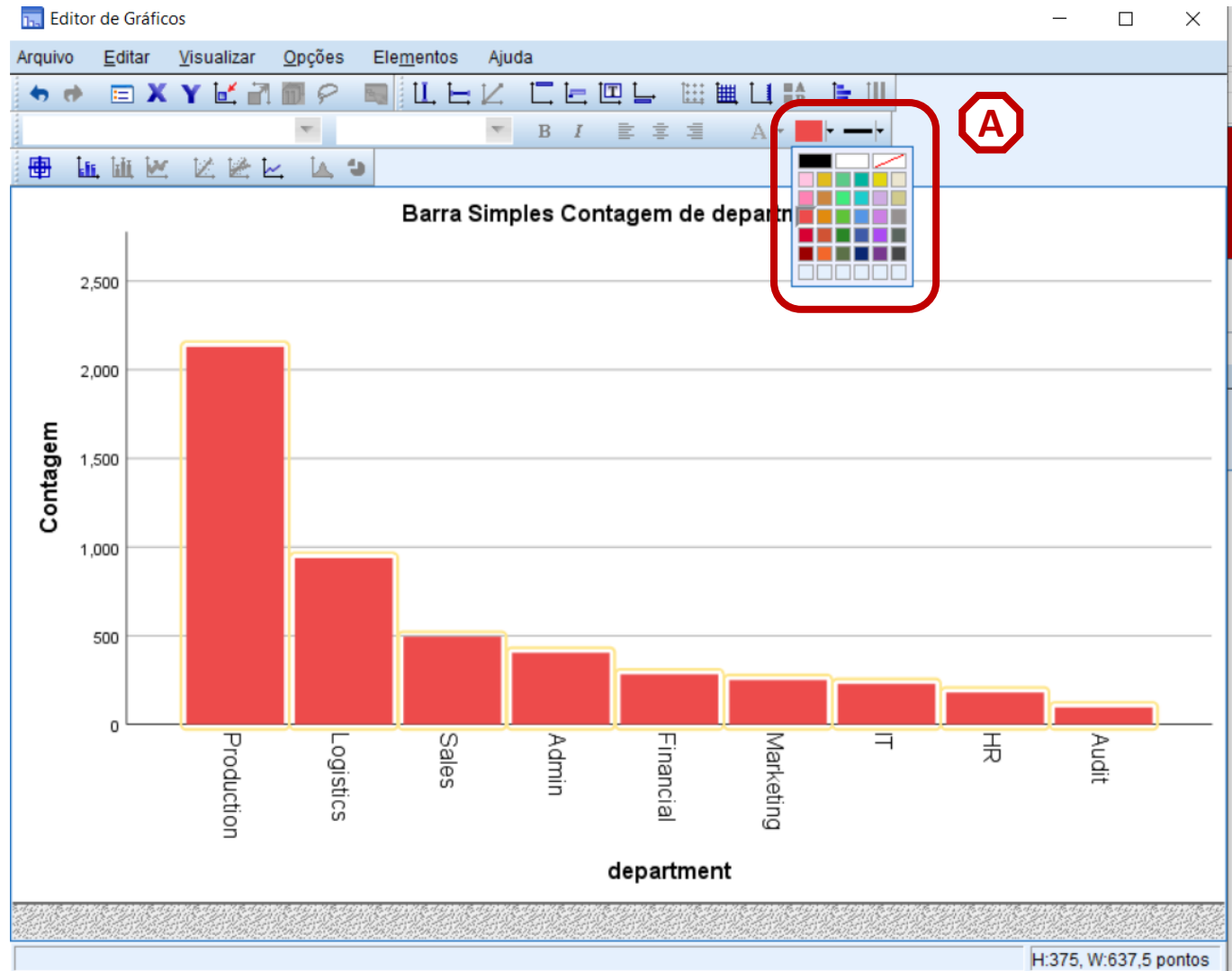


Gráfico de Barras

- 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



Análise Exploratória dos Dados

Medidas de Dispersão

Distribuição de Frequências

- **Objectivo:**
 - **Perceber se as avaliações dos trabalhadores são bastante próximas entre si, ou se há um grau elevado de dispersão?**

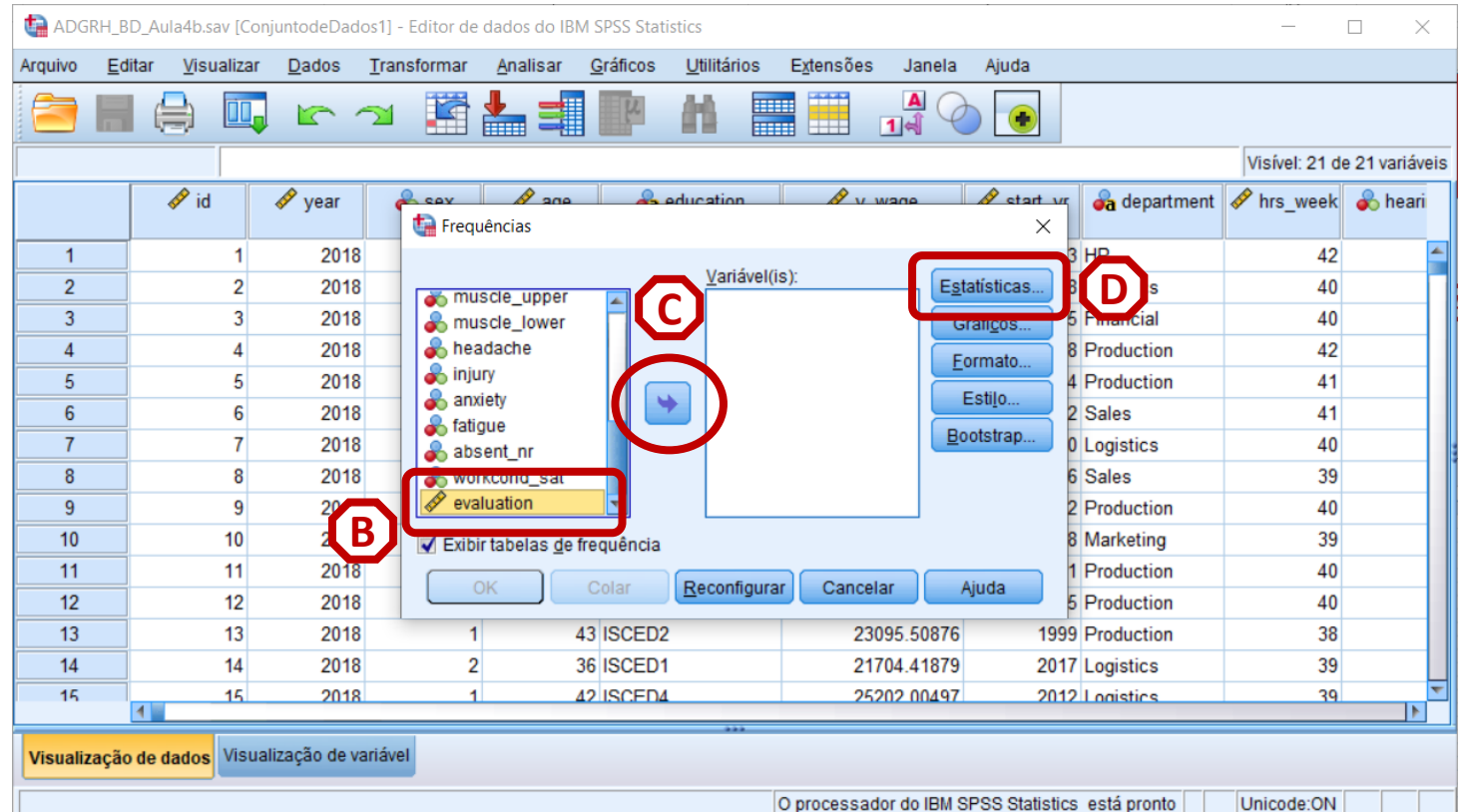
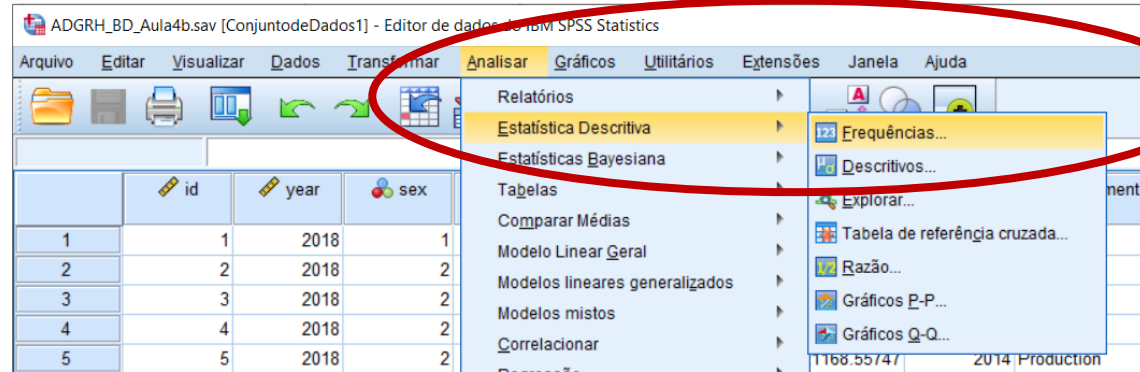
Olhar para distância entre valor máximo e valor mínimo

Olhar para a forma como se distribuem as avaliações entre decis/quartis

Olhar para as medidas de Variância e do Desvio-Padrão

Medidas de Dispersão

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



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

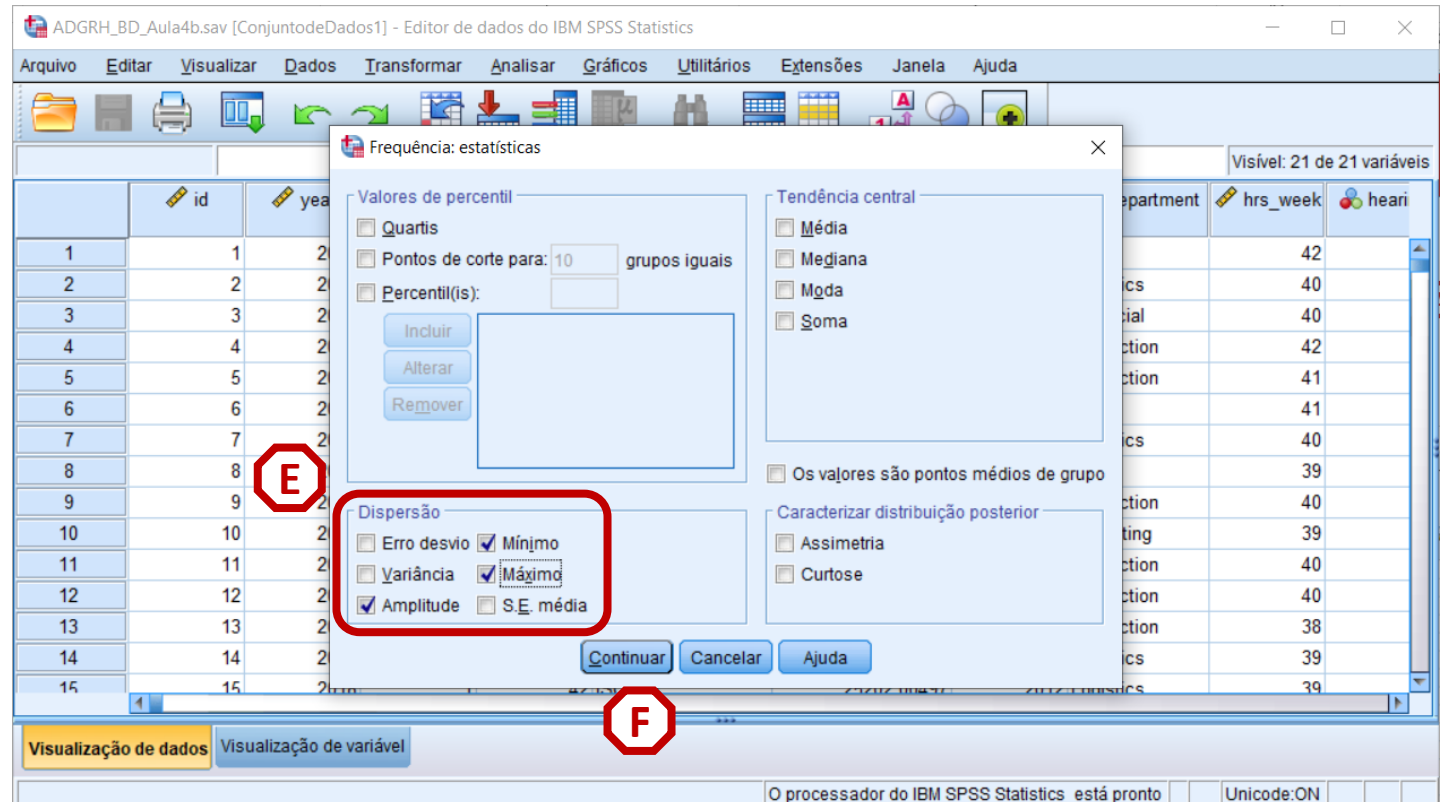
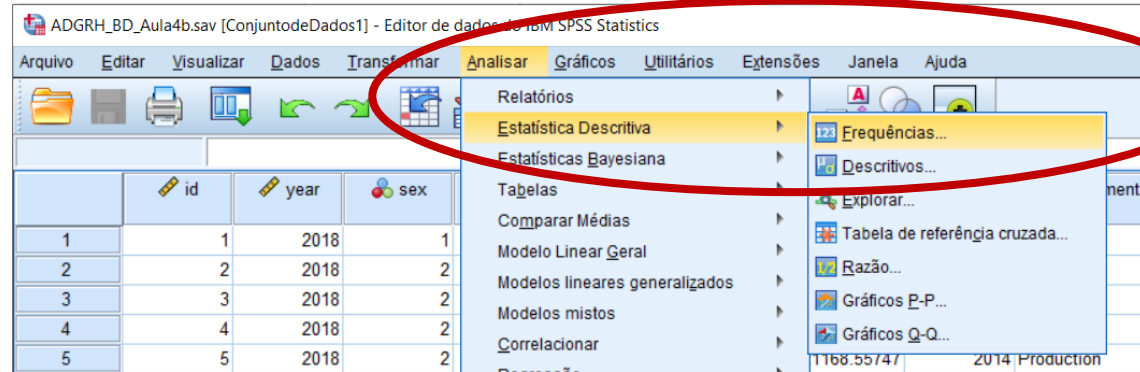
B

C

D

E

F



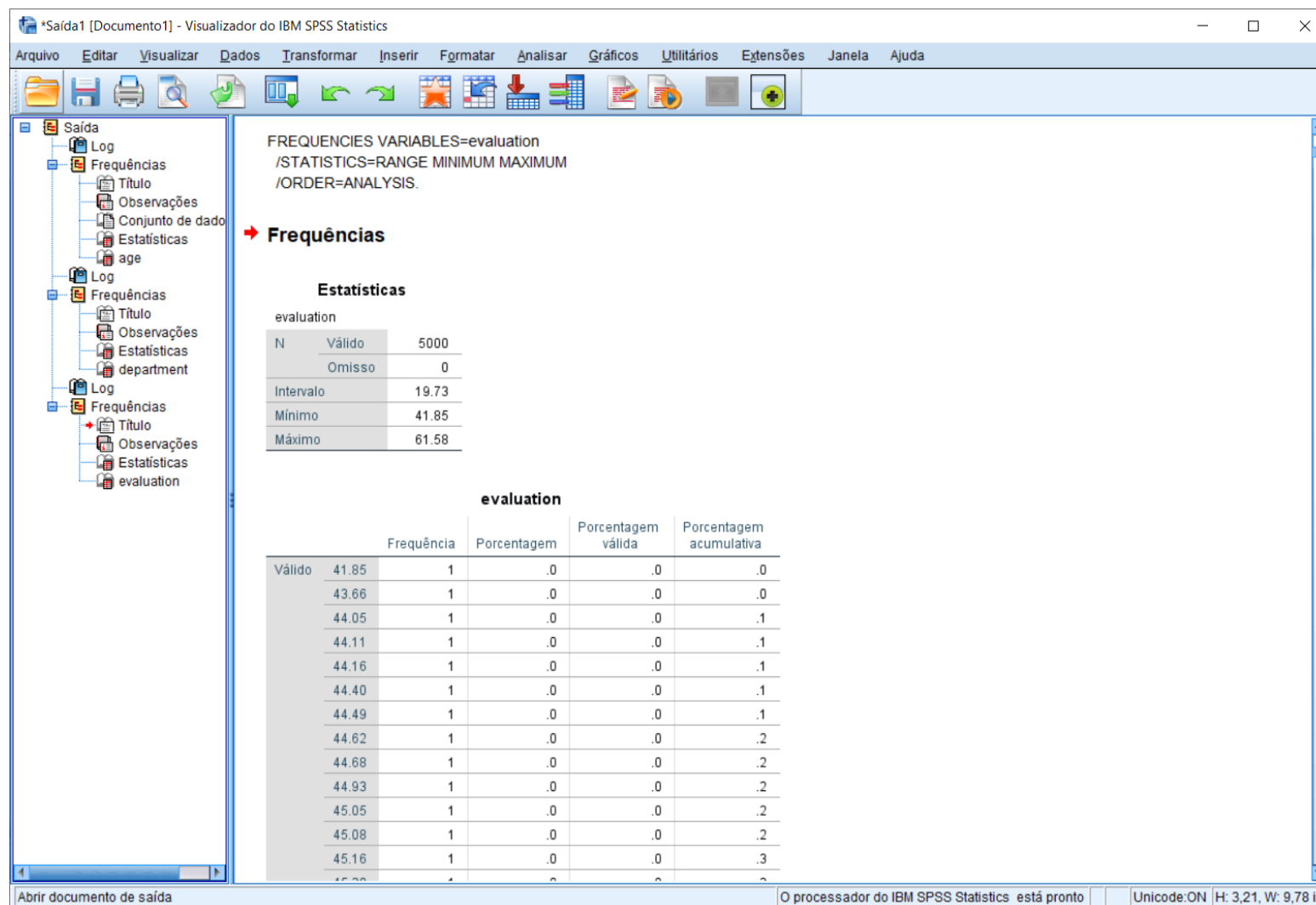
Medidas de Dispersão

- O resultado é publicado no 'Visualizador de Resultados'

Mínimo: 41.85

Máximo: 61.58

Intervalo: 19.73



The screenshot shows the IBM SPSS Statistics 'Visualizador do IBM SPSS Statistics' window. The left sidebar displays a tree view of the output, with 'Frequências' selected under the 'evaluation' variable. The main window displays the following statistics:

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

→ **Frequências**

Estatísticas

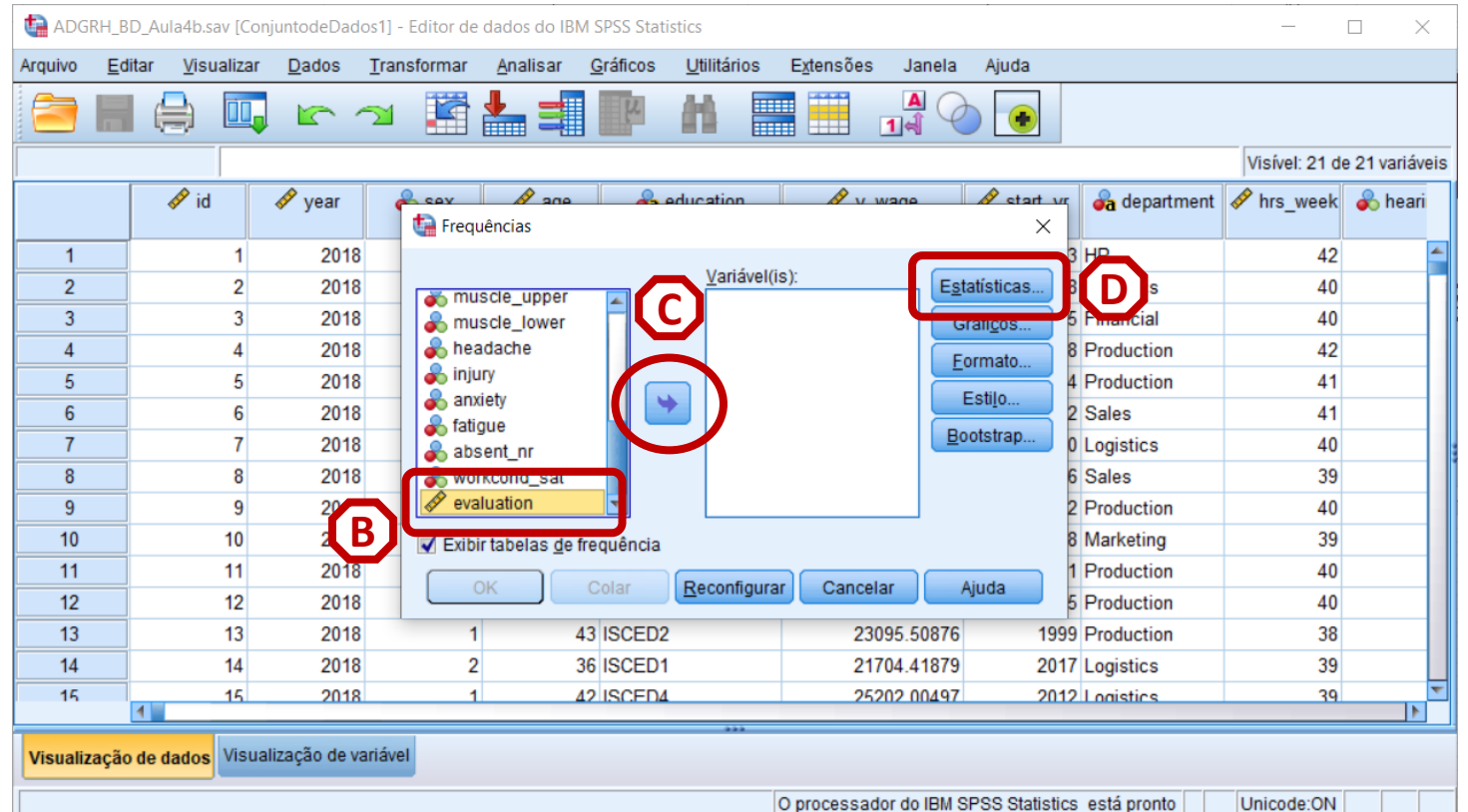
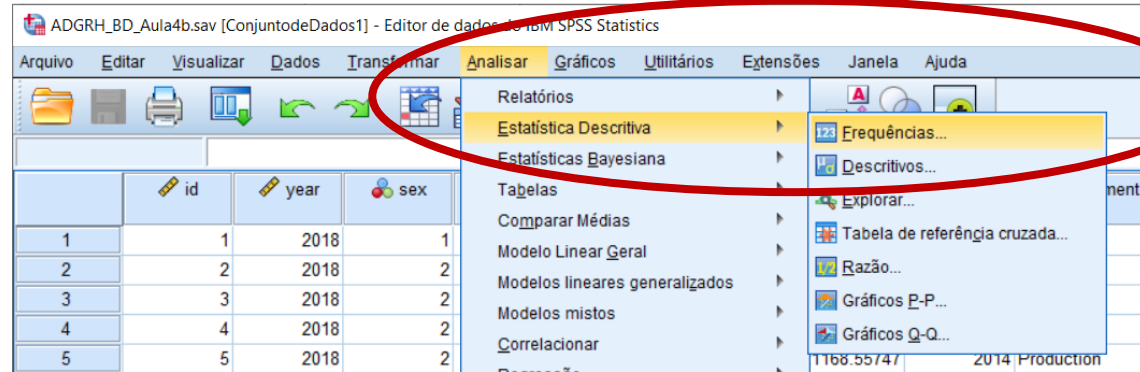
evaluation		
N	Válido	5000
	Omisso	0
Intervalo		19.73
Mínimo		41.85
Máximo		61.58

Below the summary statistics, a detailed frequency table is shown:

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

Medidas de Dispersão

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



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

ADGRH_BD_Aula4b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Visível: 21 de 21 variáveis

department hrs_week heari

1 2 2
2 2 2
3 3 2
4 4 2
5 5 2
6 6 2
7 7 2
8 8 2
9 9 2
10 10 2
11 11 2
12 12 2
13 13 2
14 14 2
15 15 2

Frequência: estatísticas

Valores de percentil

- Quartis
- Pontos de corte para: 10 grupos iguais
- Percentil(is):

Incluir Alterar Remover

Dispersão

- Erro desvio
- Mínimo
- Variância
- Máximo
- Amplitude
- S.E. média

Tendência central

- Média
- Mediana
- Moda
- Soma

Os valores são pontos médios de grupo

Caracterizar distribuição posterior

- Assimetria
- Curtose

Continuar Cancelar Ajuda

Visualização de dados Visualização de variável

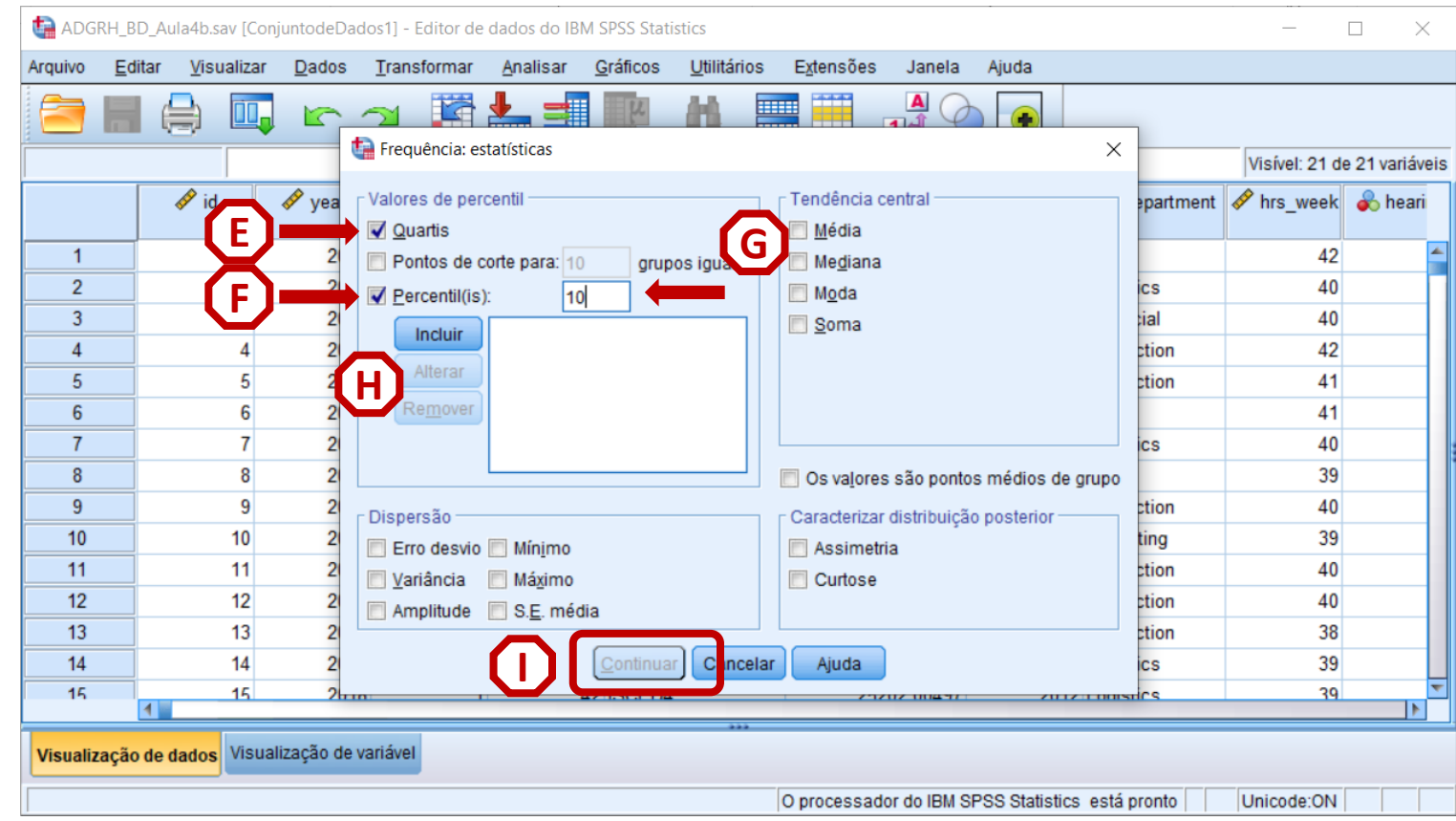
O processador do IBM SPSS Statistics está pronto Unicode:ON

Medidas de Dispersão

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

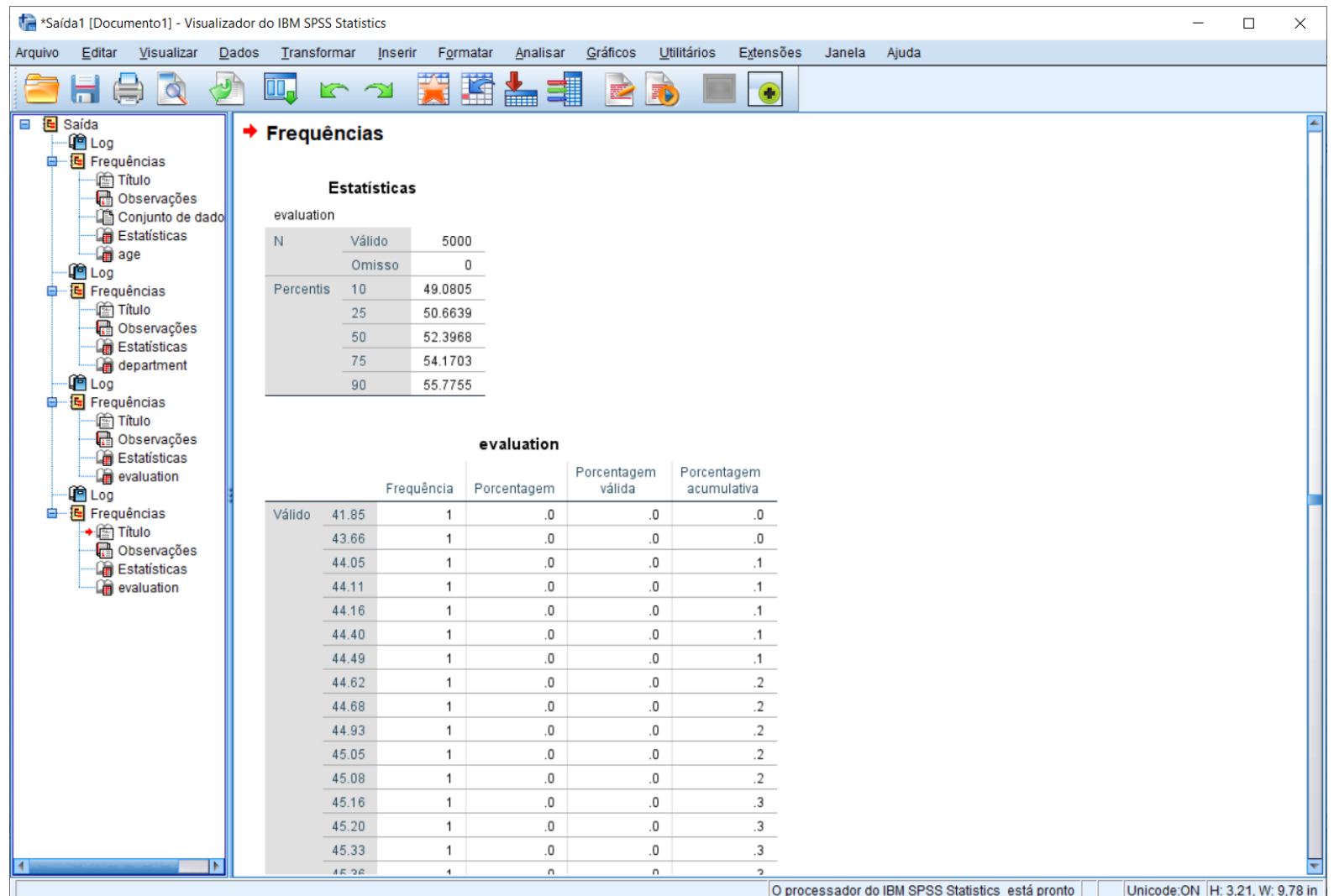
Exercício: Inserir 9º Decil

- Seleccionar 'Continuar' / 'OK'



Medidas de Dispersão

- O resultado é publicado no ‘Visualizador de Resultados’
- 1º Decil = 49.08
- 1º Quartil = 50.66
- 2º Quartil = 52.39
- 3º Quartil = 54.17
- 9º Decil = 55.77



The screenshot shows the 'Visualizador do IBM SPSS Statistics' window. The left sidebar displays a tree view of the output, with 'Frequências' selected for the 'evaluation' variable. The main area shows the 'Estadísticas' table and a detailed frequency table.

Estadísticas

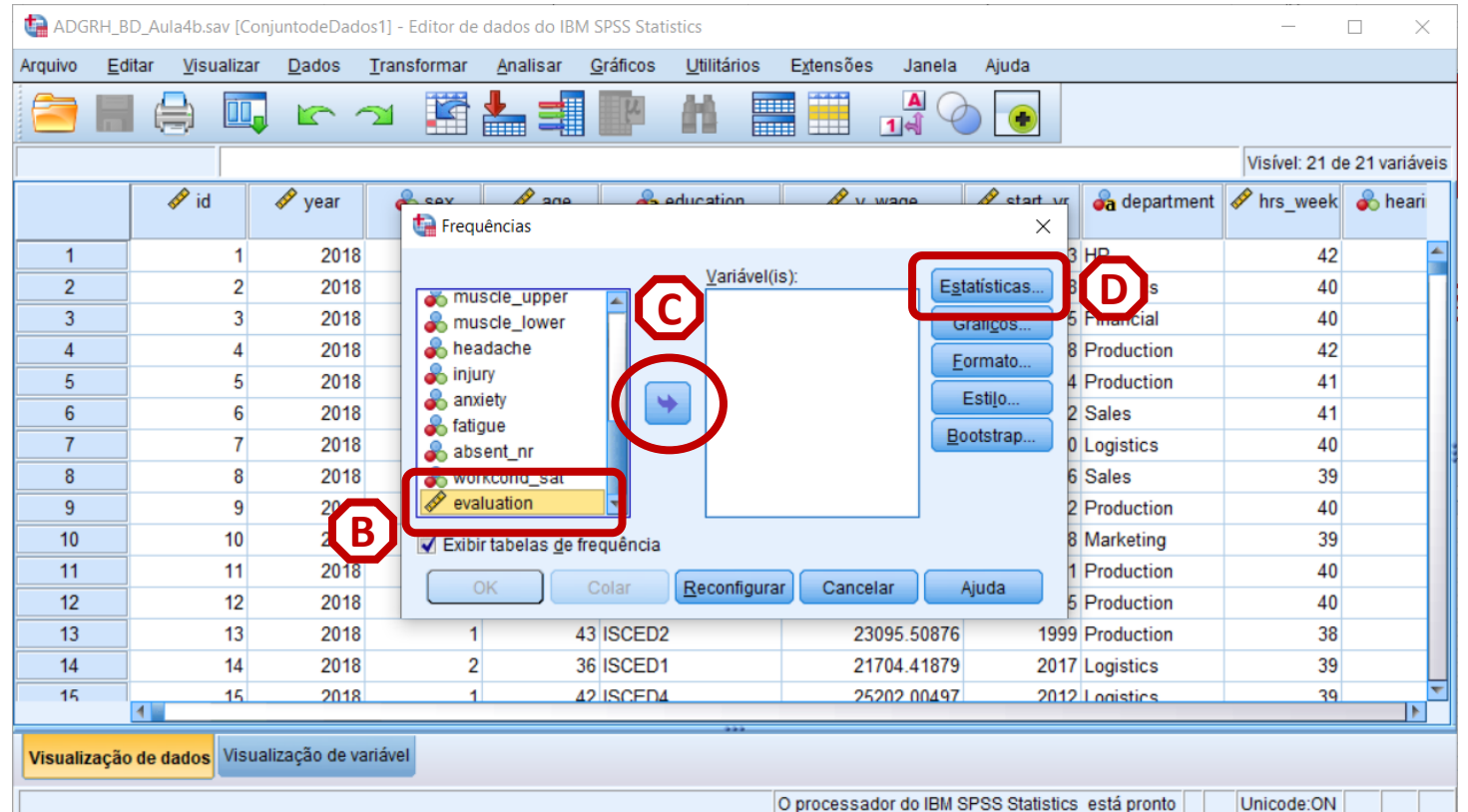
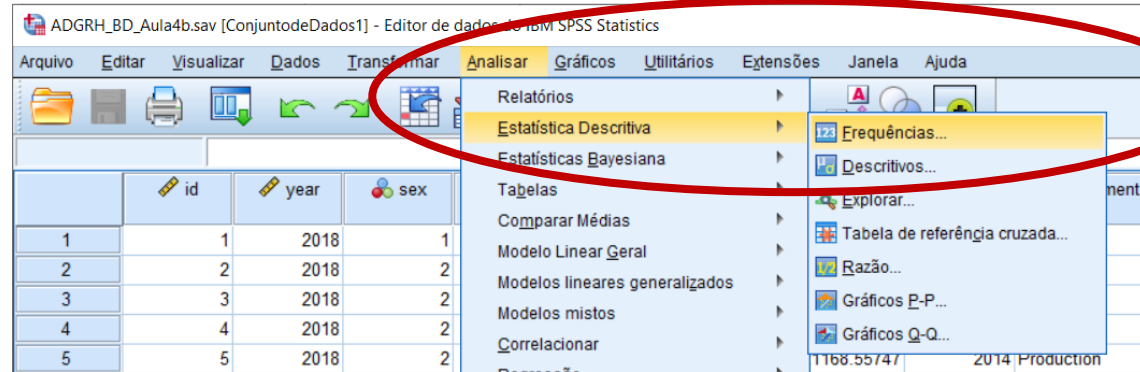
evaluation		
N	Válido	5000
	Omisso	0
Percentis	10	49.0805
	25	50.6639
	50	52.3968
	75	54.1703
	90	55.7755

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
	45.20	1	.0	.3
	45.33	1	.0	.3
	45.26	1	.0	.2

Medidas de Dispersão

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



Medidas de Dispersão

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Frequências' A
- Selecionar variável 'evaluation' B
- Colocar na caixa de variáveis C
- Selecionar 'Estatísticas' D
- Selecionar 'Erro Desvio' (Desvio Padrão) 'Variância' E
- Selecionar 'Média' F
- Selecionar 'Continuar' / 'OK' G

A

B

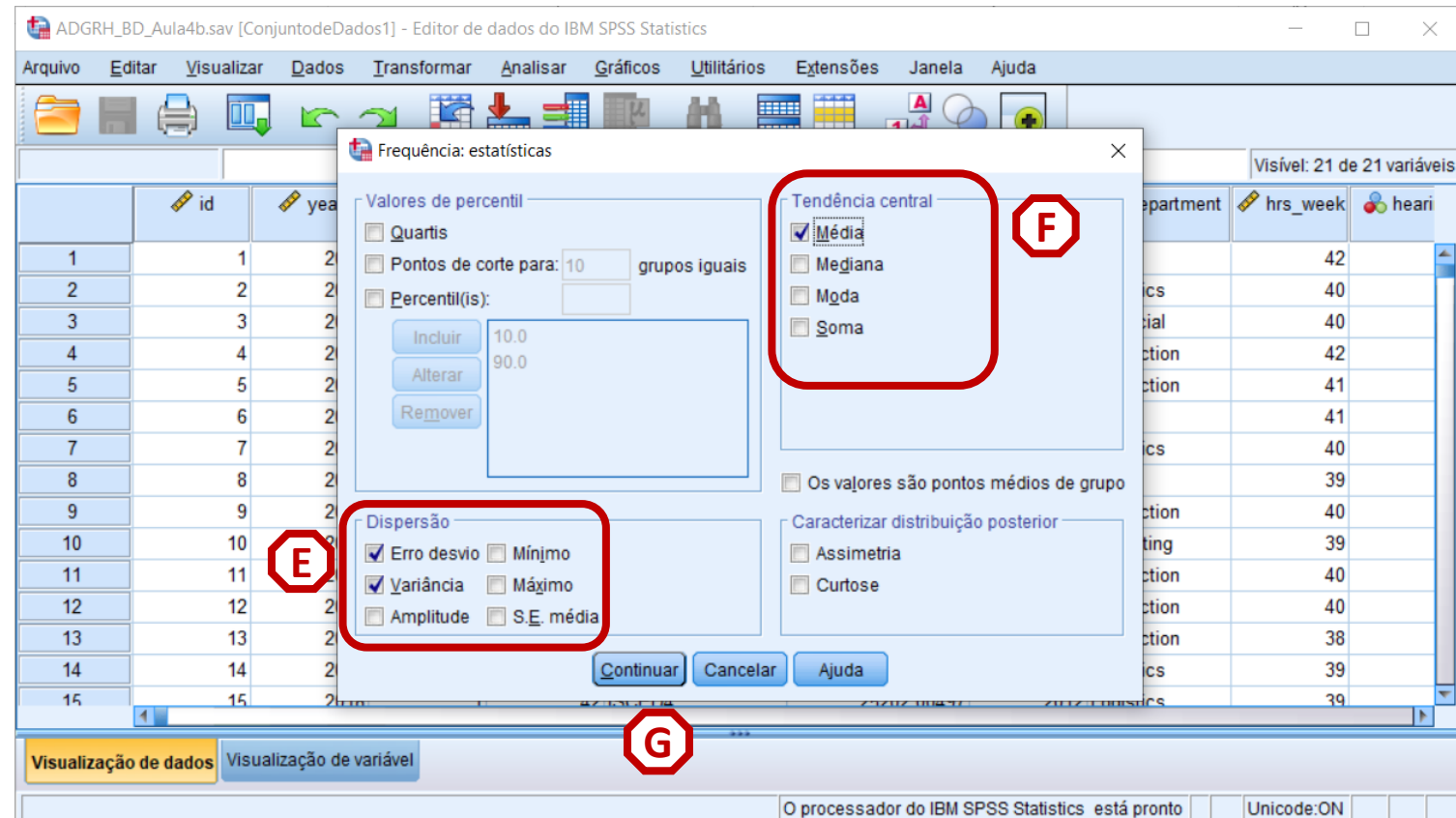
C

D

E

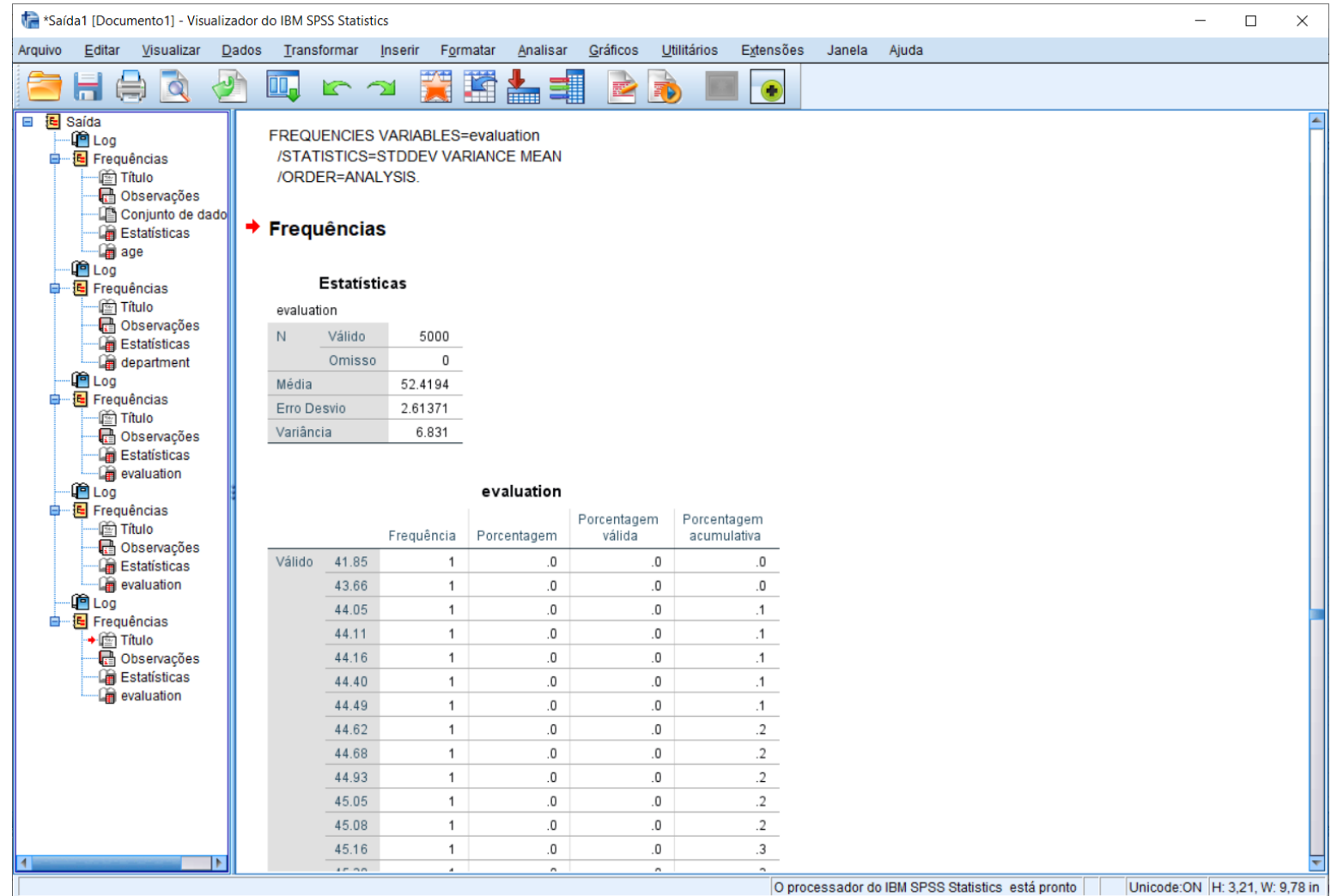
F

G



Medidas de Dispersão

- O resultado é publicado no 'Visualizador de Resultados'
- Média = 52.41
- Desvio Padrão = 2.61
- Variância = 6.83



FREQUENCIES VARIABLES=evaluation
/STATISTICS=STDDEV VARIANCE MEAN
/ORDER=ANALYSIS.

→ **Frequências**

Estatísticas

evaluation		
N	Válido	5000
	Omisso	0
Média		52.4194
Erro Desvio		2.61371
Variância		6.831

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

O processador do IBM SPSS Statistics está pronto | Unicode.ON | H: 3,21, W: 9,78 in

Medidas de Dispersão

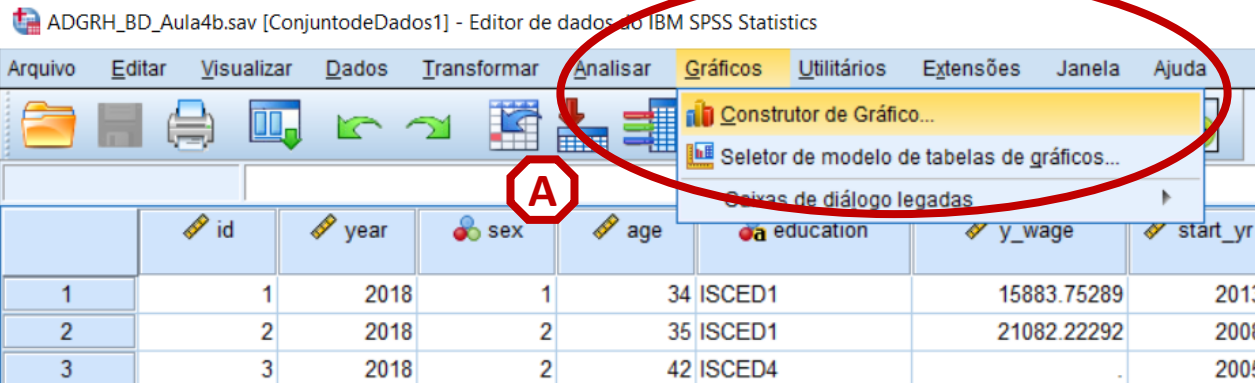
- **QUAL É A MELHOR FORMA DE VISUALIZAR O GRAU DE DISPERSÃO?**

GRÁFICO DE CAIXA (*E BIGODES*)

Gráfico de Caixa

- Selecionar 'Gráficos' / 'Construtor de Gráfico'

A



The screenshot shows the IBM SPSS Statistics interface. The 'Gráficos' menu is open, and the 'Construtor de Gráfico...' option is highlighted. A red circle highlights the 'Gráficos' menu and the 'Construtor de Gráfico...' option. A red octagon with the letter 'A' is placed over the 'Gráficos' menu. Below the menu, a data table is visible with columns: id, year, sex, age, education, y_wage, and start_yr.

	id	year	sex	age	education	y_wage	start_yr
1	1	2018	1	34	ISCED1	15883.75289	2013
2	2	2018	2	35	ISCED1	21082.22292	2008
3	3	2018	2	42	ISCED4	.	2005

Gráfico de Caixa

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

A

B

C

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- id
- year
- sex
- age
- education
- y_wage
- start_yr
- department
- hrs_week
- hearino

Nenhuma categoria (variável de escala)

Boxplot em 1-D

Eixo X?

Filtro?

Galeria

Escolha entre:

- Favoritos
- Barra
- Linha
- Área
- Setor/Polar
- Dispersão/Ponto
- Histograma
- 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...

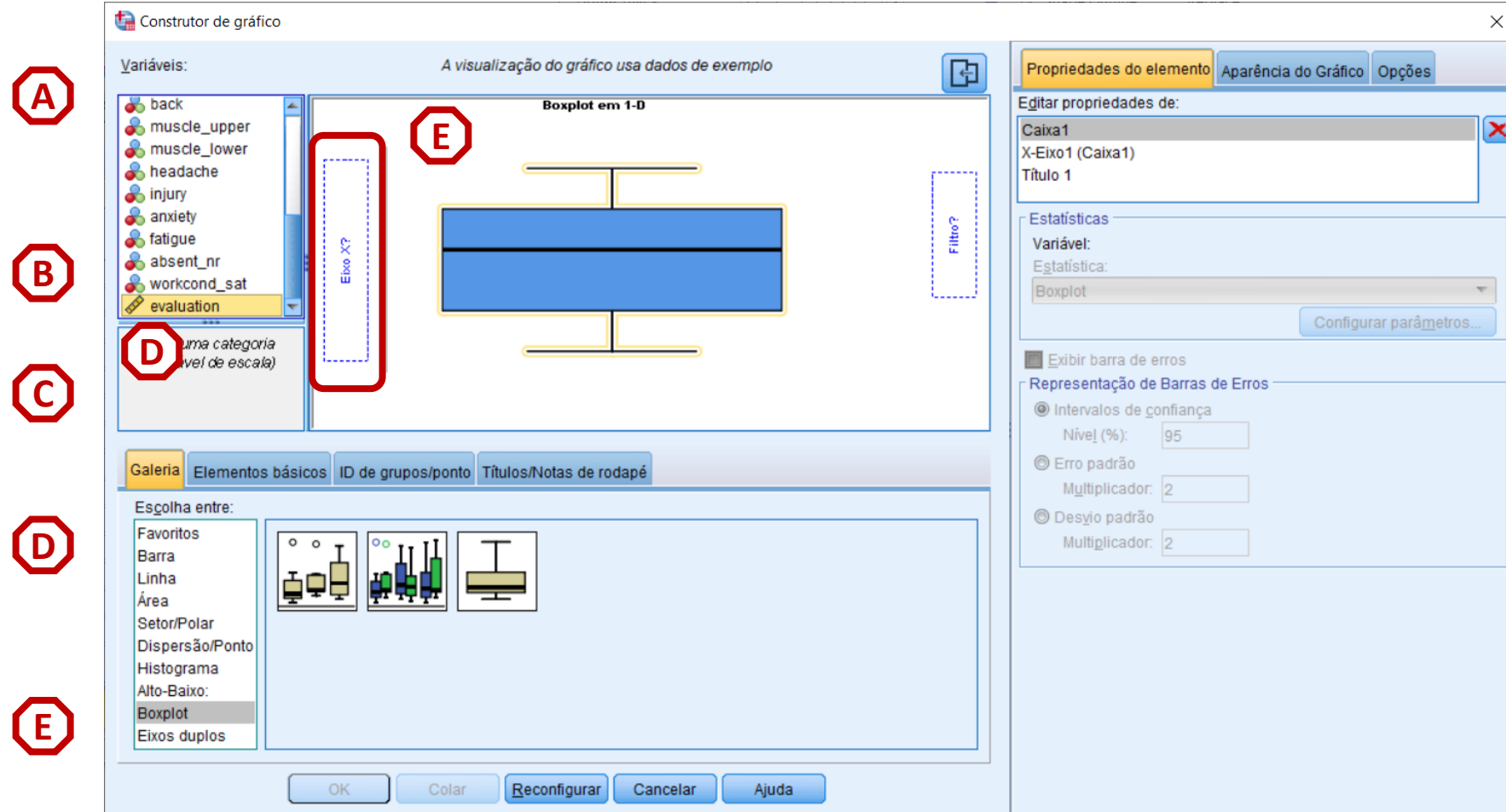
Exibir barra de erros

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'



The screenshot shows the 'Construtor de gráfico' window with a 'Boxplot em 1-D' visualization. The interface includes a 'Variáveis:' list on the left, a central plot area, and a 'Galeria' (Gallery) at the bottom. Red octagons A-E highlight the following steps:

- A**: The 'Construtor de gráfico' window title.
- B**: The 'Boxplot' option selected in the 'Galeria'.
- C**: The 'evaluation' variable selected in the 'Variáveis:' list.
- D**: The 'Eixo X' label in the plot area.
- E**: The 'Eixo X?' label in the plot area.

The right-hand 'Propriedades do elemento' panel shows the configuration for 'Caixa1', including 'Estatísticas' and 'Representação de Barras de Erros'.

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

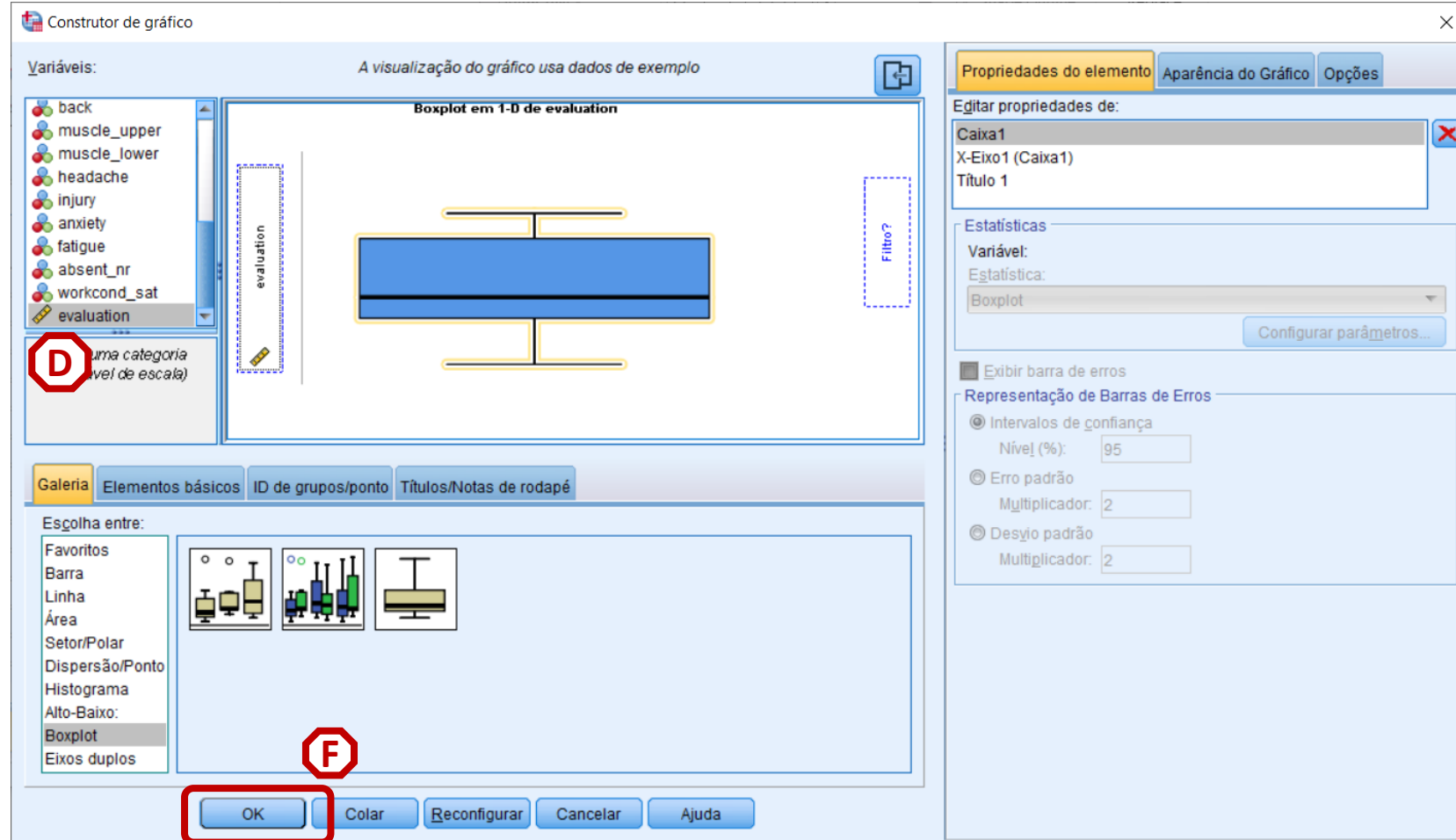
B

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 (vel de escala)

Boxplot em 1-D de evaluation

OK Colar Reconfigurar Cancelar Ajuda

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de:

Caixa1

X-Eixo1 (Caixa1)

Título 1

Estadísticas

Variável:

Estadística:

Boxplot

Configurar parâmetros...

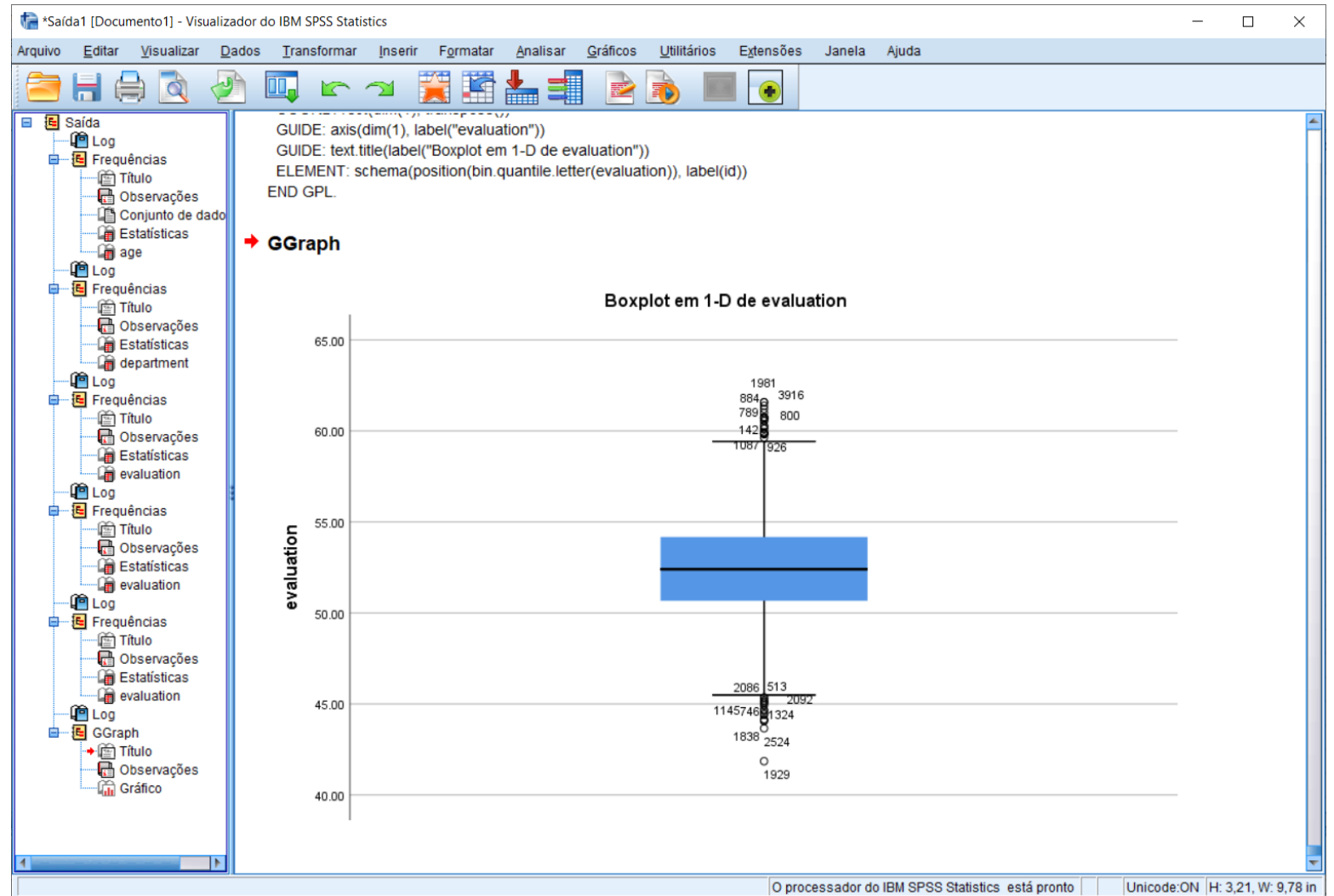
Exibir barra de erros

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'



Limpar a base de dados para o SPSS

Identificar/Codificar 'Valores Extremos'
Codificar Valores Ausentes

Limpar a base de dados para o SPSS

Identificar/Codificar Valores Extremos

- **O que são valores extremos**
 - **Em termos estatísticos: São casos/observações que estão muito distantes da (curva de) distribuição normal.**
 - **Em termos práticos:**
 - **Casos que - por erro na resposta, erro na imputação, ou por representarem casos verdadeiramente atípicos – se diferenciam significativamente dos restantes casos na amostra;**
 - **Podemos ter valores extremos na parte de cima da distribuição, mas também na parte de baixo;**
- **Por que é que são importantes**
 - **Por se afastarem da distribuição normal vão enviesar o cálculo de todo um conjunto de medidas estatísticas (ver acima)**

- **Como identificar valores extremos**
 - Inspeção Visual (Gráfico de Caixa)
 - Abordagem qualitativa: Inspeccionar casos para decidir se fazem sentido
 - Usando medidas de dispersão (z-scores, desvio padrão)

- **Como devemos tratar estes casos**
 - Abordagem mais simples: recodificar como valor ausente
 - Recodificar como valor mais (alto/baixo) – Só em casos que justifiquem!!
 - Abordagem mais avançada: métodos de clusterização

Visualizar valores extremos

- Seleccionar 'Gráficos' / 'Construtor de Gráfico'



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

Arquivo Editar Visualizar Dados Transformar Analisar **Gráficos** Utilitários Extensões Janela Ajuda

Construtor de Gráfico...
Seletor de modelo de tabelas de gráficos...

Caixas de diálogo legadas

	year	sex	age	education	start_y	start_age	experie e	y_wage
1	2018.00	1.00	32	ISCED4	2016.00	30.00	2.00	22153.05
2	2018.00	1.00	40	ISCED4	2003.00	25.00	15.00	27809.35



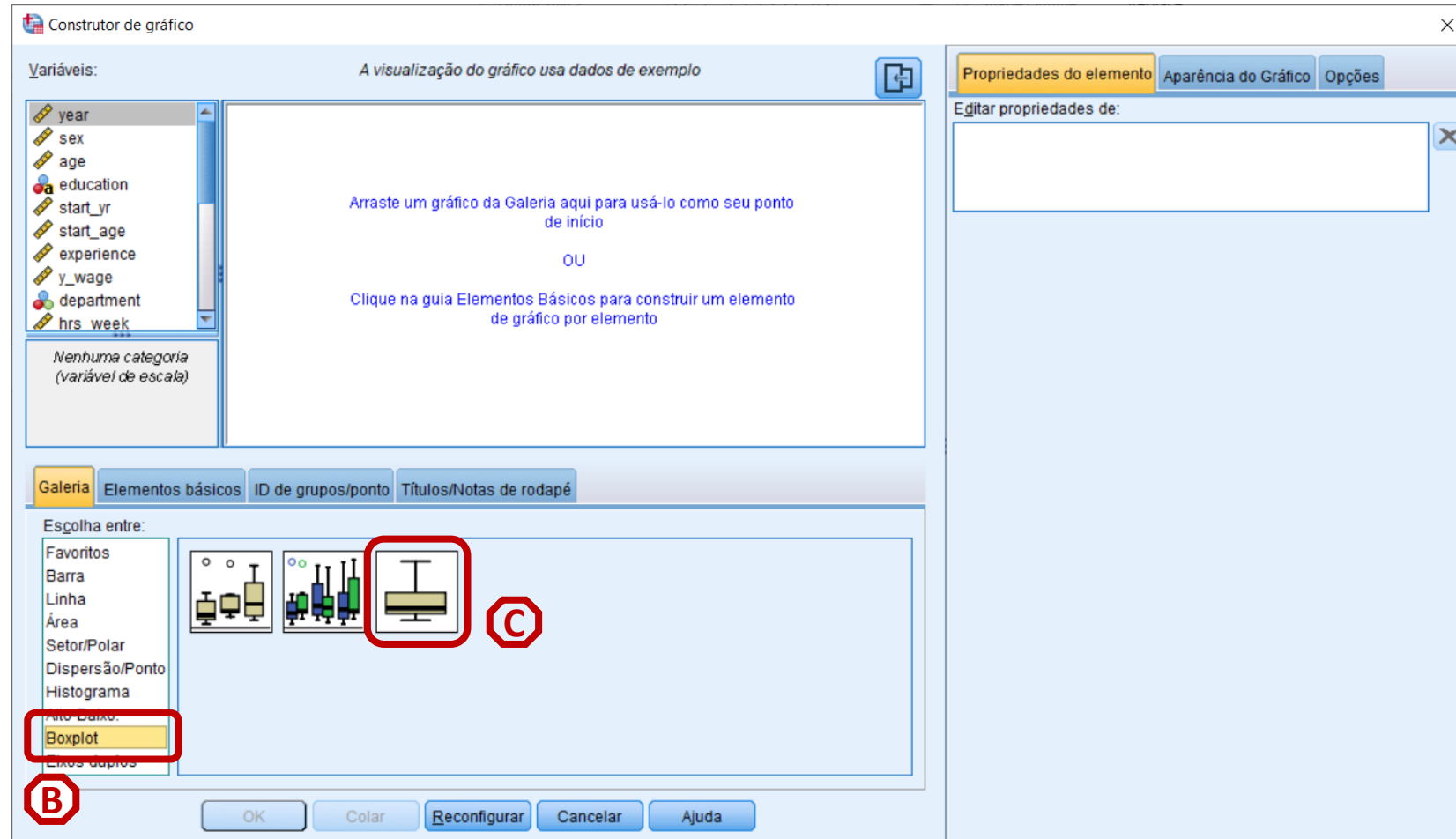
Visualizar valores extremos

- Seleccionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Diagrama em Caixa'
- Duplo-Clique no Gráfico de Caixa (simples)

A

B

C



Visualizar valores extremos

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Diagrama em Caixa'
- Duplo-Clique no Gráfico de Caixa (simples)
- Selecionar a variável 'y_wage'

A
B
C
D

The screenshot shows the 'Construtor de gráfico' window. The 'Variáveis:' list on the left contains 'year', 'sex', 'age', 'education', 'yr', 'age', 'y_wage', 'department', and 'hrs week'. 'y_wage' is highlighted with a red box labeled 'B'. A red circle labeled 'A' is around the 'Gráficos' / 'Construtor de Gráfico' button. A red circle labeled 'C' is around the 'Boxplot' option in the 'Escolha entre:' gallery. A red circle labeled 'D' is around the 'y_wage' variable in the list. The main visualization area shows a 'Boxplot em 1-D' with a blue box and whiskers. The right panel shows the 'Propriedades do elemento' for 'Caixa1' with various statistical options.

Visualizar valores extremos

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Diagrama em Caixa'
- Duplo-Clique no Gráfico de Caixa (simples)
- Selecionar a variável 'y_wage'
- Colocar a variável 'y_wage' no 'Eixo X'
- Selecionar 'OK'

A

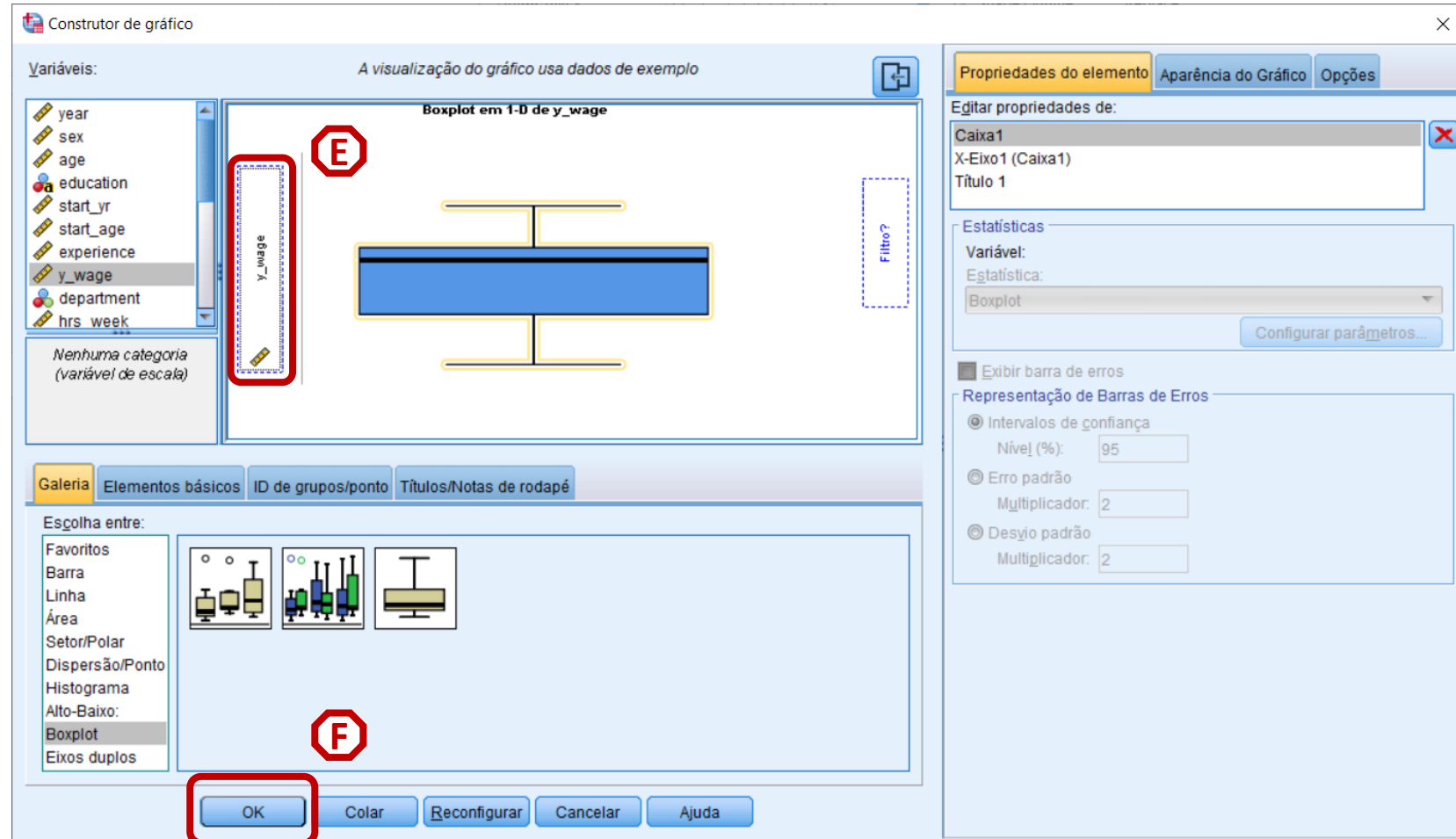
B

C

D

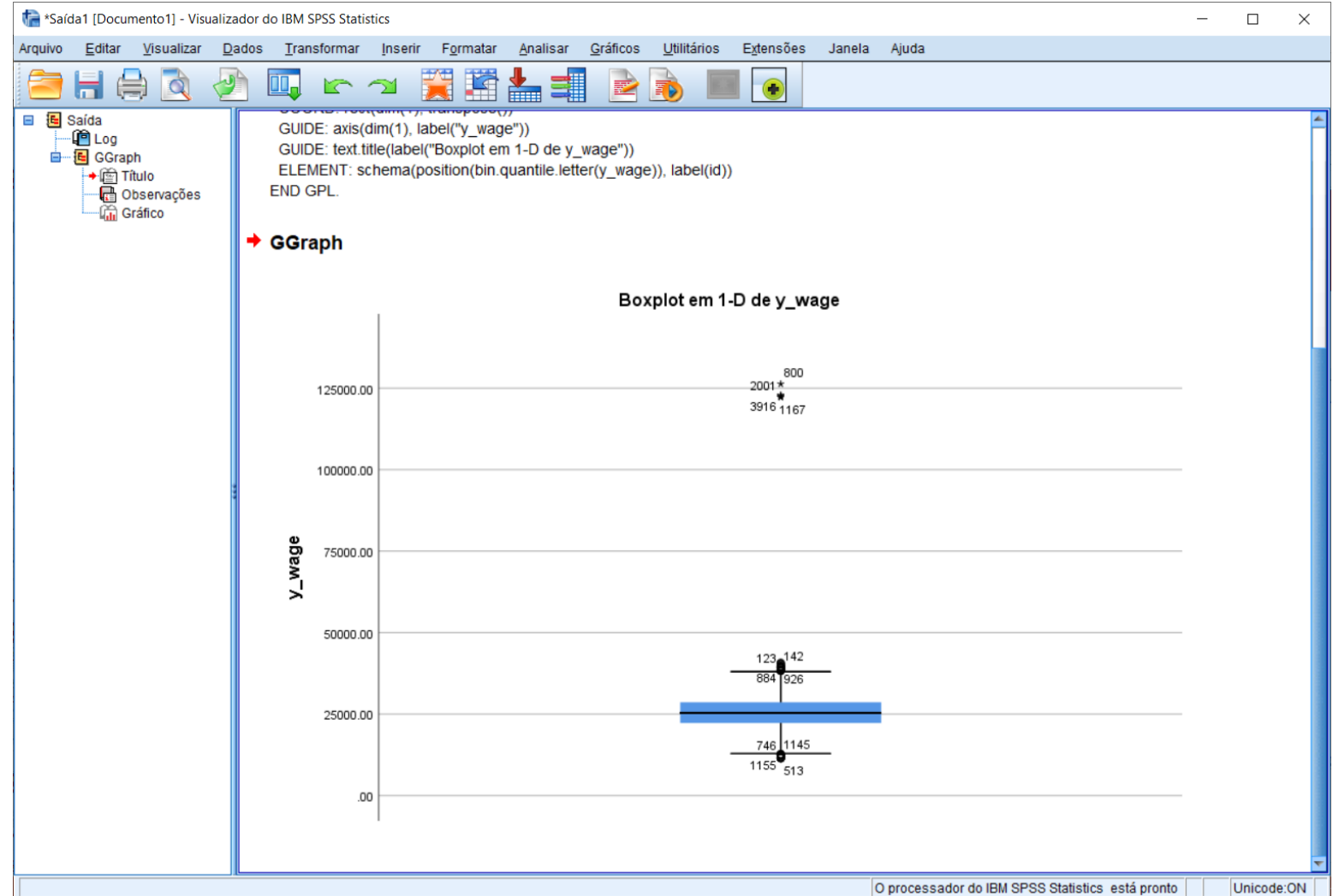
E

F



Visualizar valores extremos

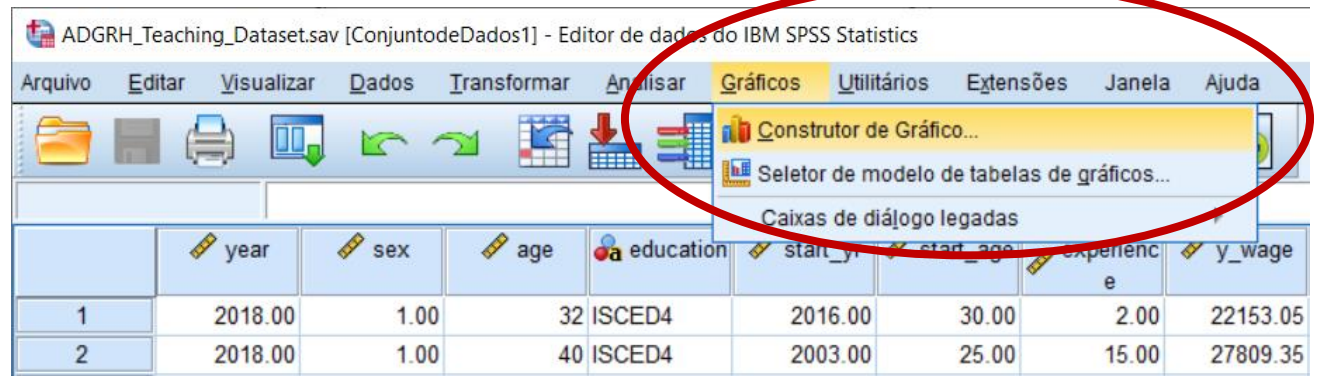
- O gráfico é publicado no 'Visualizador de Resultados'
- Neste caso, poderíamos recodificar todos os valores acima de 50.000.



- **Mas nem sempre os dados extremos são tão fáceis de identificar...**

Visualizar valores extremos

- Seleccionar 'Gráficos' / 'Construtor de Gráfico'



Visualizar valores extremos

- Seleccionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Diagrama em Caixa'
- Duplo-Clique no Gráfico de Caixa (simples)

A

B

C

Visualizar valores extremos

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Diagrama em Caixa'
- Duplo-Clique no Gráfico de Caixa (simples)
- Selecionar a variável 'y_wage2'

A

B

C

D

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- muscle_upper
- muscle_lower
- headache
- injury
- anxiety
- fatigue
- ent_nr
- cond_sat
- evaluation
- y_wage2**

Nenhuma categoria (variável de escala)

Boxplot em 1-D

Eixo X?

Filtro?

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de:

Caixa 1

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

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

Visualizar valores extremos

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Diagrama em Caixa'
- Duplo-Clique no Gráfico de Caixa (simples)
- Selecionar a variável 'y_wage'
- Colocar a variável 'y_wage' no 'Eixo X'
- Selecionar 'OK'

A

B

C

D

E

F

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- muscle_upper
- muscle_lower
- headache
- injury
- anxiety
- fatigue
- absent_nr
- workcond_sat
- evaluation
- y_wage2

Nenhuma categoria (variável de escala)

Boxplot em 1-D de y_wage2

Y-Eixo 1 (Caixa1)

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...

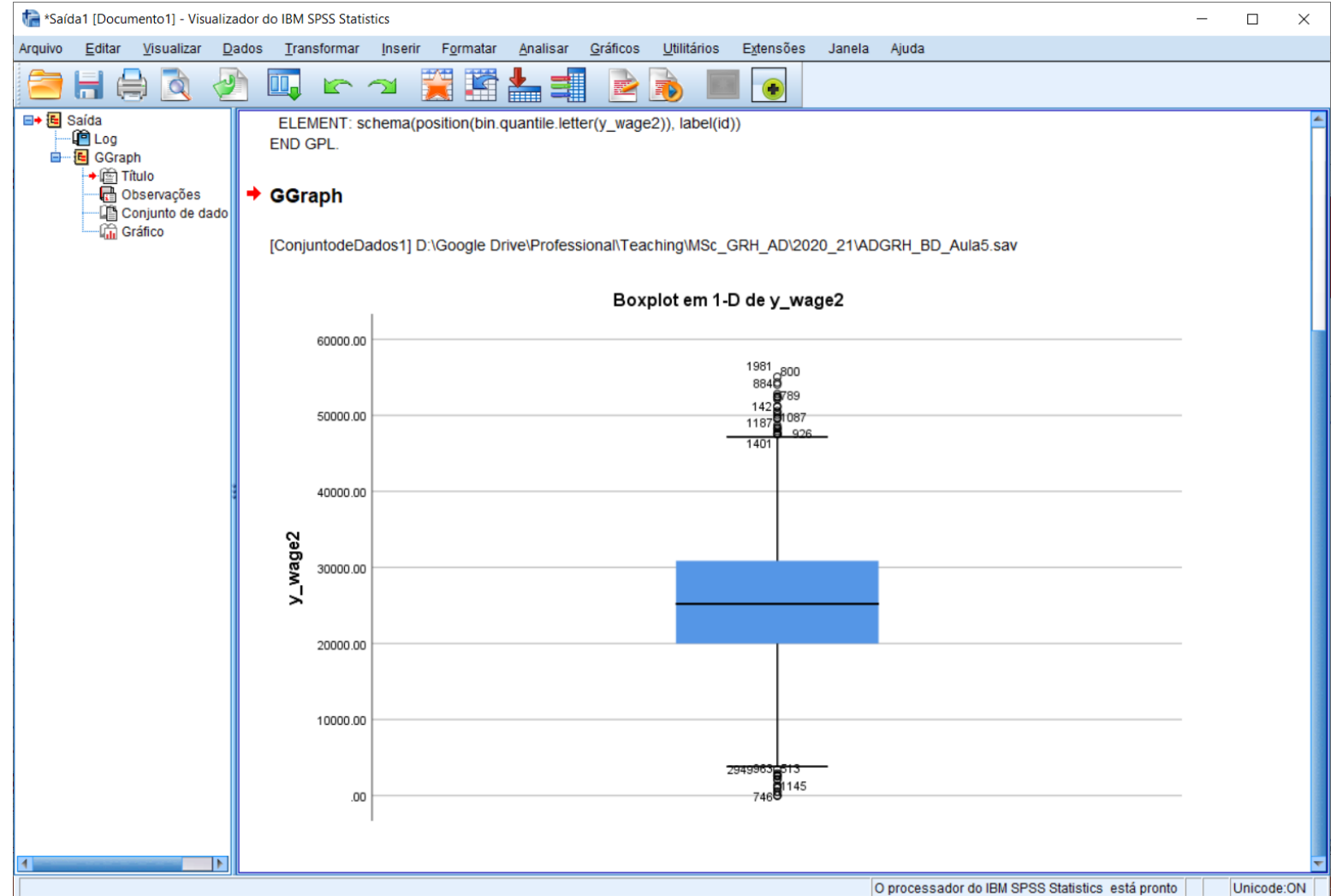
Exibir barra de erros

Representação de Barras de Erros

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

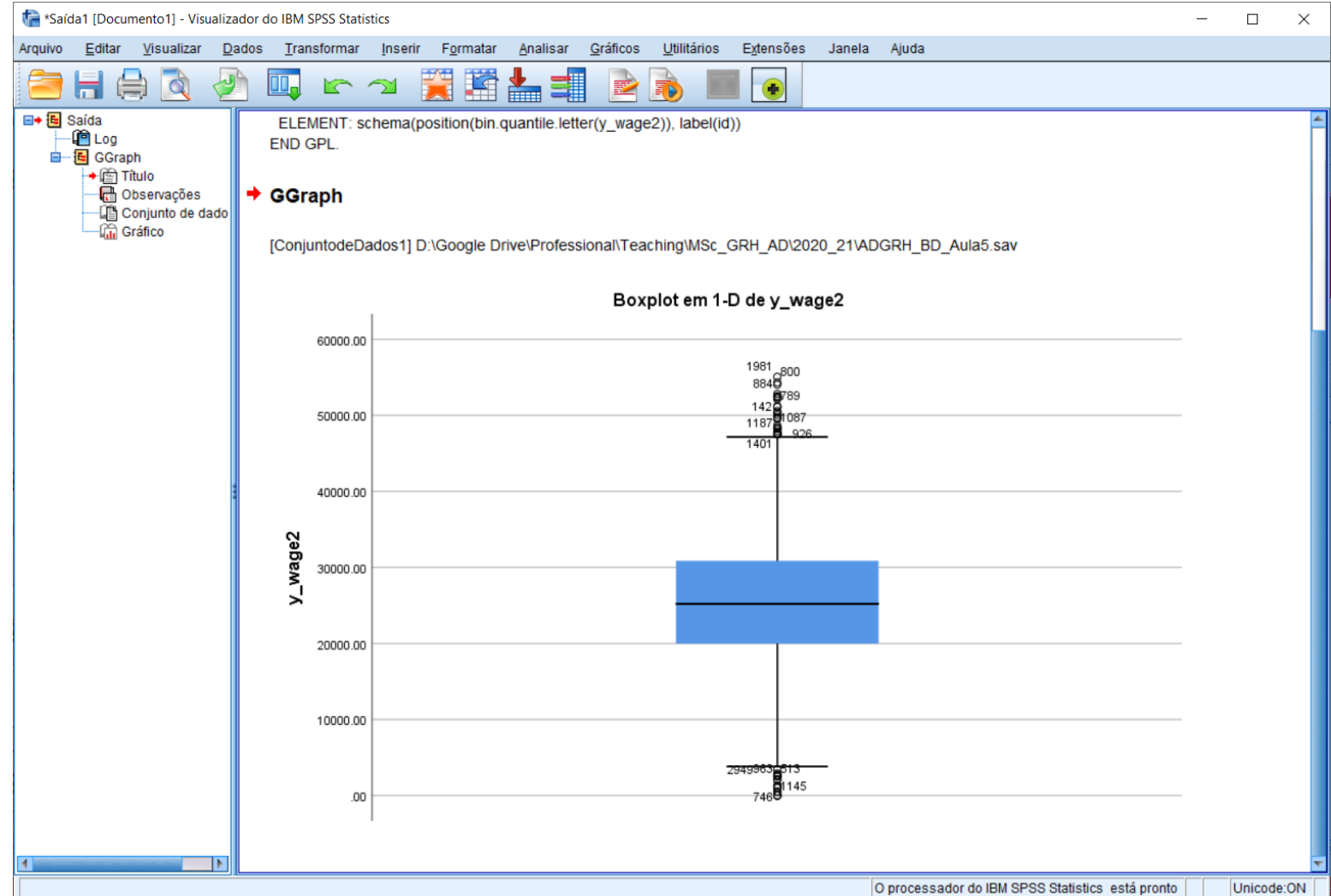
Visualizar valores extremos

- O gráfico é publicado no 'Visualizador de Resultados'
- Assim, à primeira vista, alguém consegue identificar outliers nesta variável?



Visualizar valores extremos

- O gráfico é publicado no 'Visualizador de Resultados'
- Assim, à primeira vista, alguém consegue identificar outliers nesta variável?
- Parece haver muitos valores próximos do 0!
- A não ser que o dono seja o Tio Patinhas... estes valores não fazem sentido.

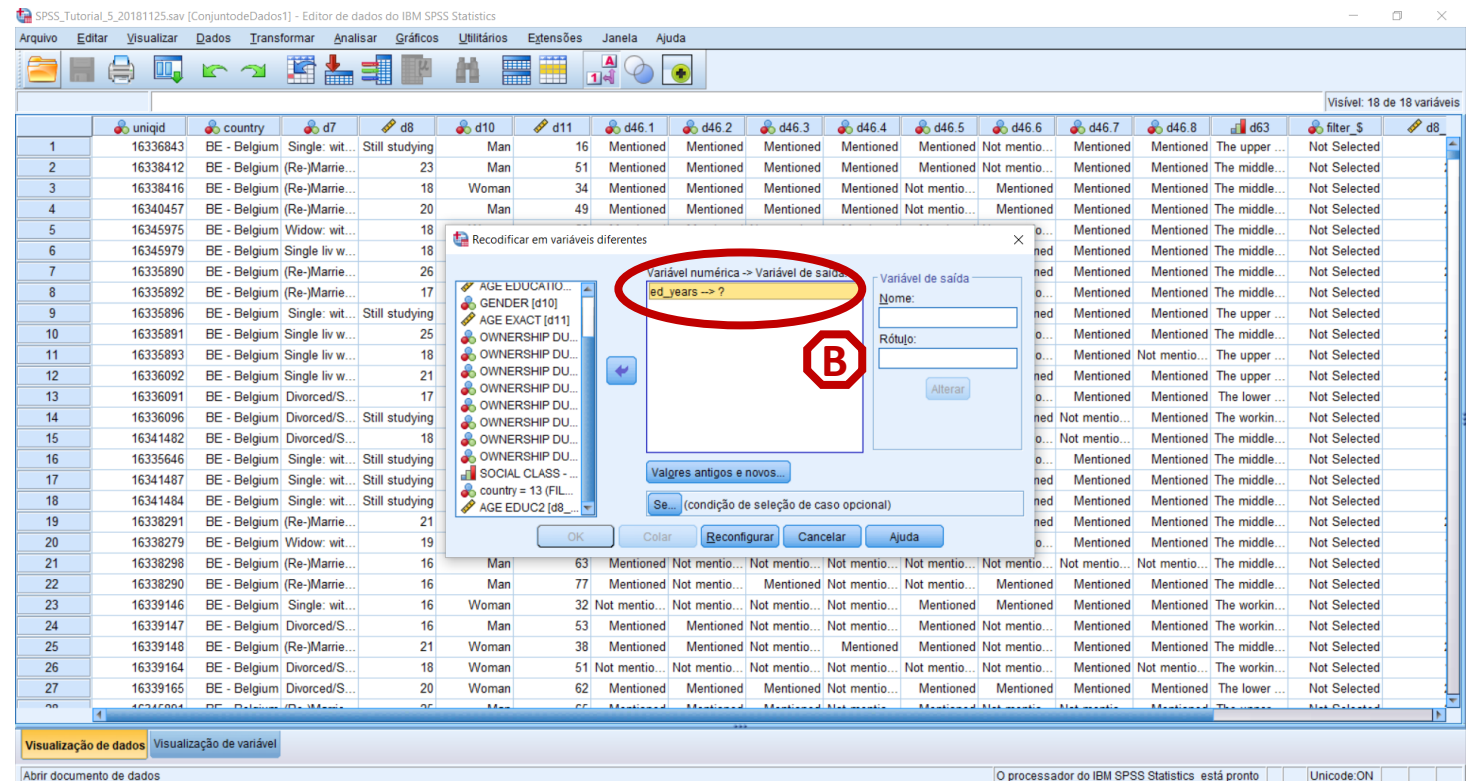
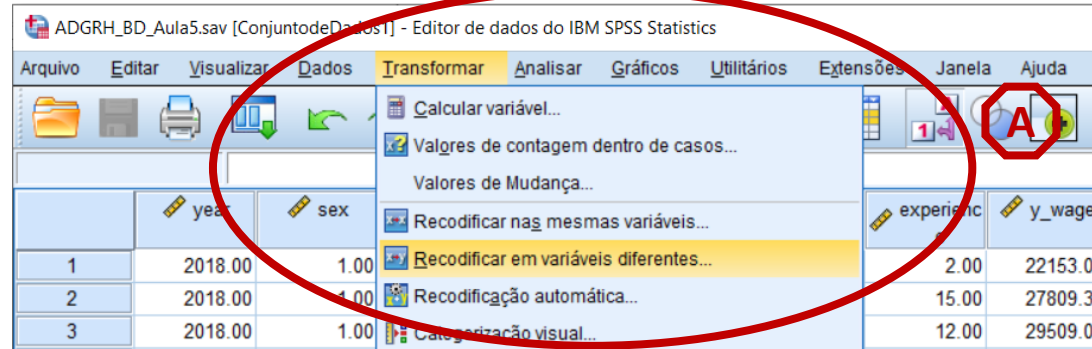


Recodificar valores extremos

- **Objectivo:**
 - **Recodificar todos os menores ou iguais a €8120 (salário mínimo nacional, anual)...**
- **... Como valores salários iguais a €8120.**

Recodificar valores extremos

- Seleccionar 'Transformar' / 'Recodificar em Variáveis Diferentes'
- Definir 'y_wage2' como Variável de Entrada



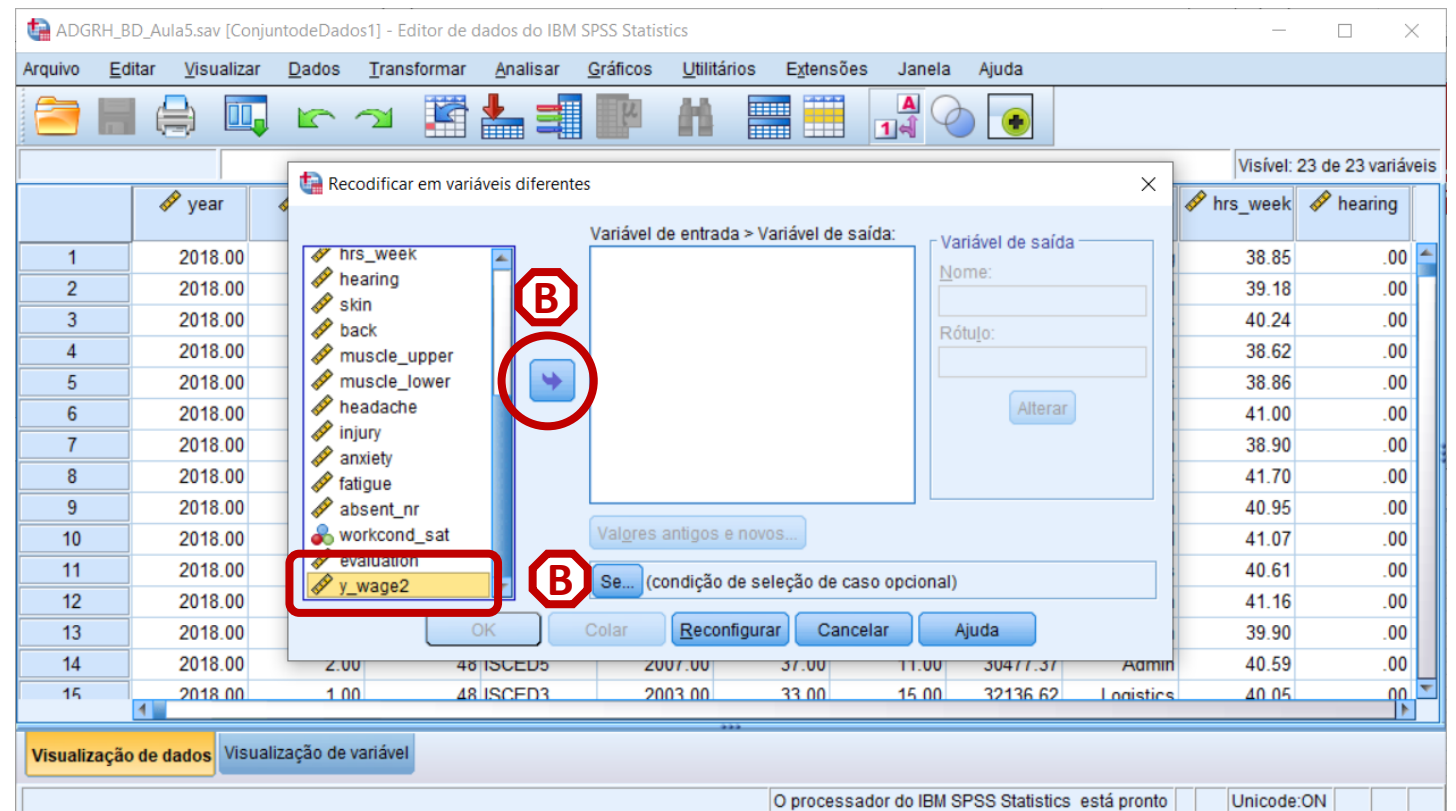
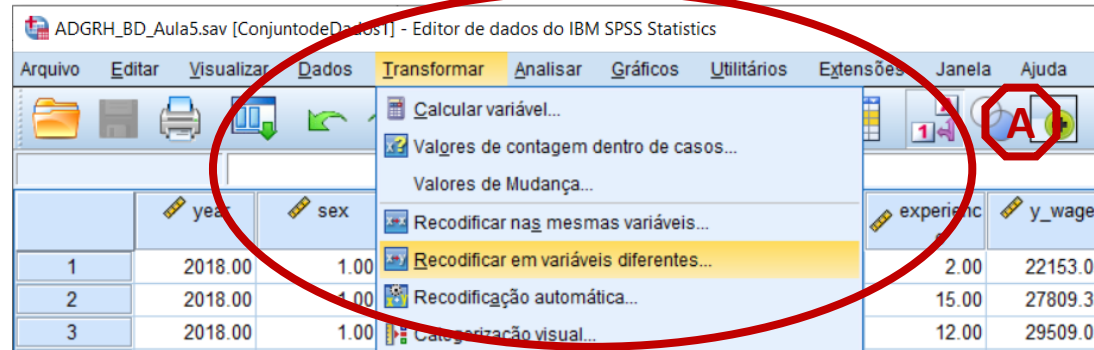
Recodificar valores extremos

- Selecionar 'Transformar' / 'Recodificar em Variáveis Diferentes'
- Selecionar a variável 'y_wage2'
- Definir 'y_wage2' como Variável de Entrada

A

B

C



Recodificar valores extremos

- Selecionar 'Transformar' / 'Recodificar em Variáveis Diferentes'
- Selecionar a variável 'y_wage2'
- Definir 'y_wage2' como Variável de Entrada
- Definir o nome da nova variável 'y_wage2'... e o rótulo 'WAGE_TMP'
- Selecionar 'Alterar'
- Selecionar 'Valores antigo e novo'

A

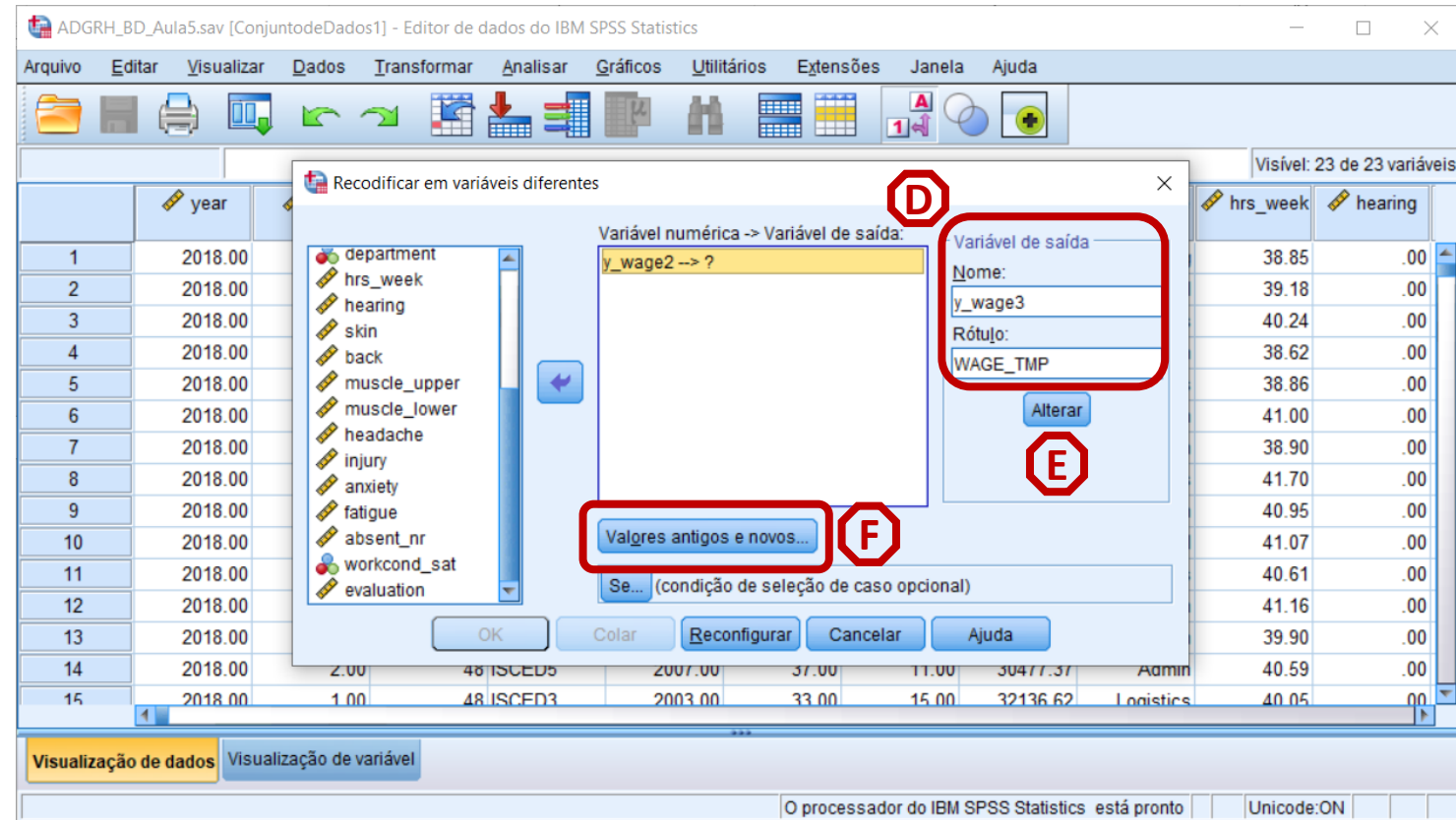
B

C

D

E

F



ADGRH_BD_Aula5.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Recodificar em variáveis diferentes

Variável numérica -> Variável de saída:
y_wage2 --> ?

Variável de saída
Nome: y_wage3
Rótulo: WAGE_TMP
Alterar

Valores antigos e novos...

Se... (condição de seleção de caso opcional)

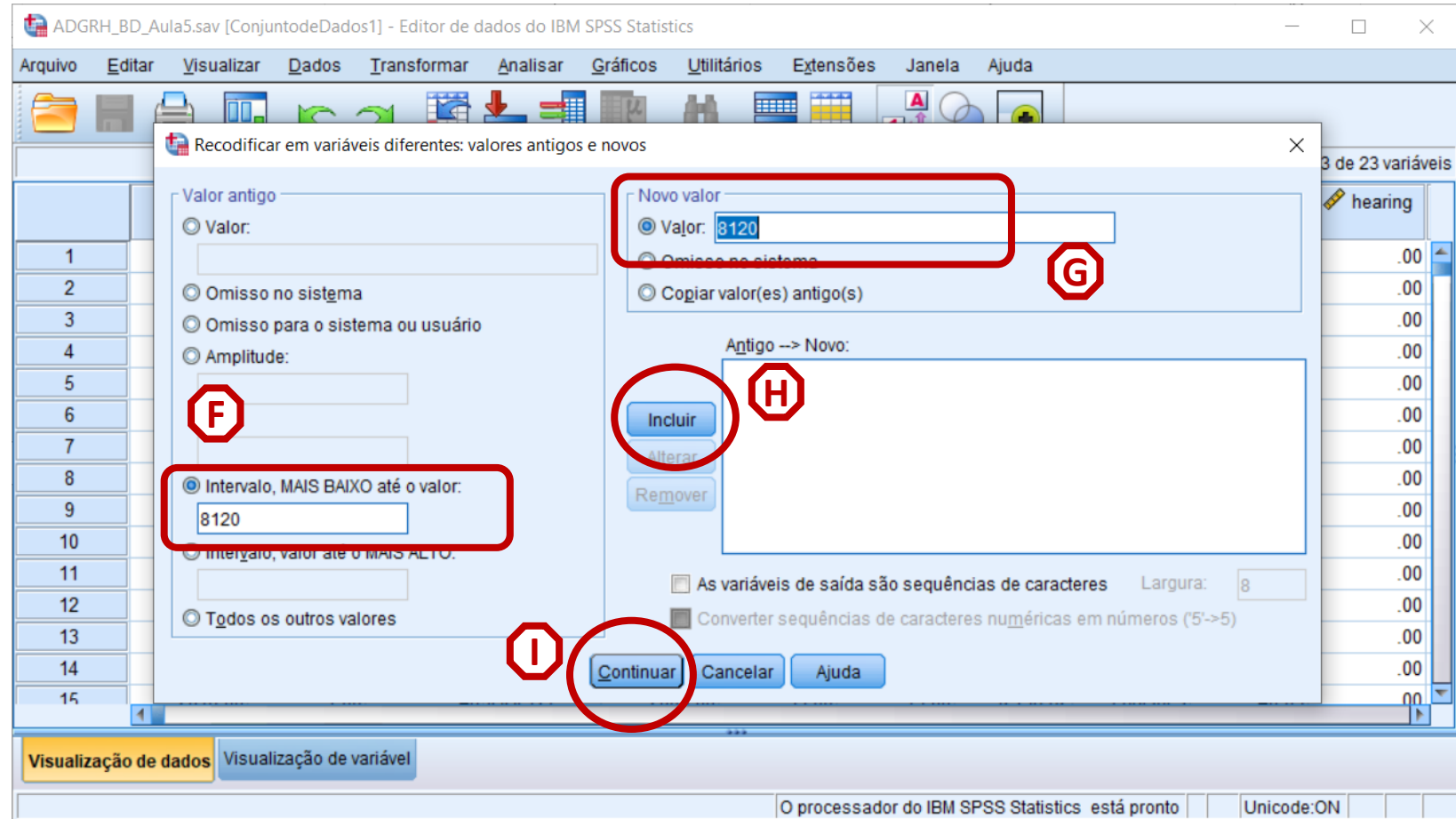
OK Colar Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

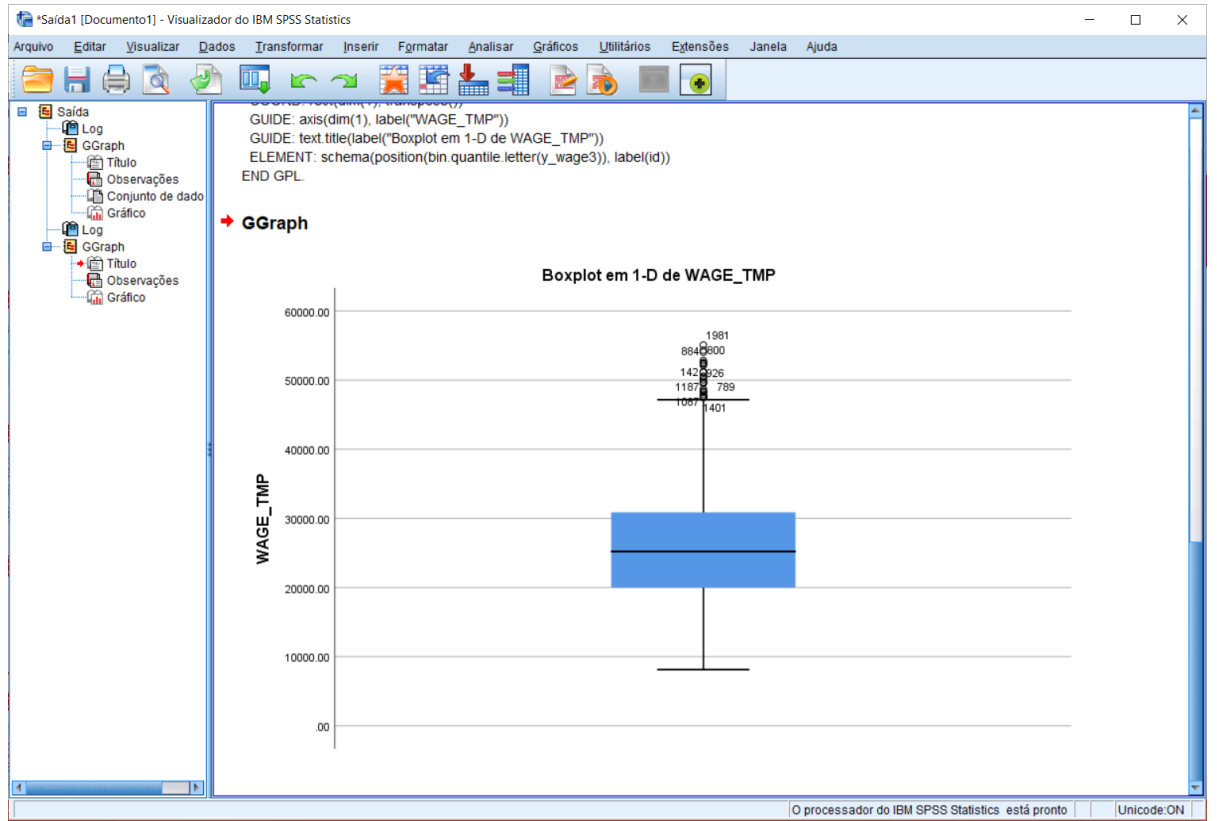
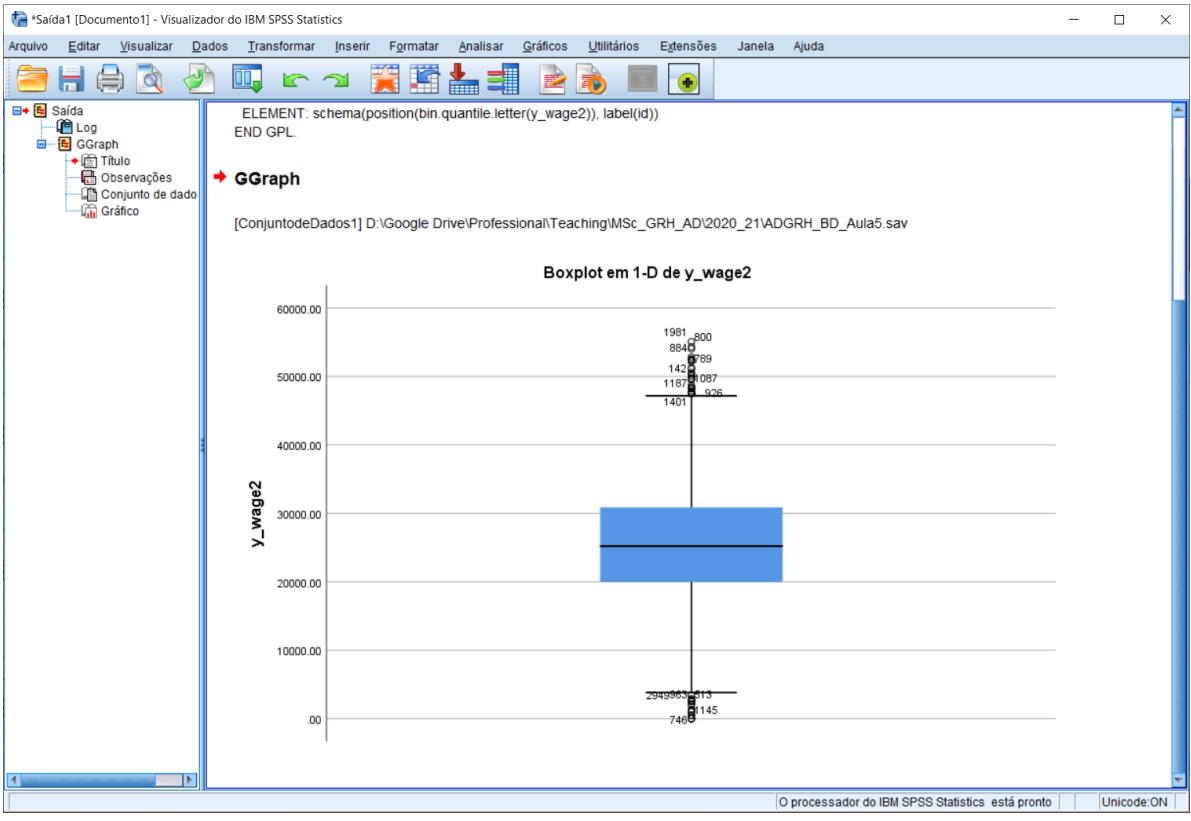
O processador do IBM SPSS Statistics está pronto Unicode:ON

Recodificar valores extremos

- Os casos ≤ 8120 na variável original
- Definidos como '8120' na nova variável
- Selecionar 'Incluir'
- Exercício:
'Todos os outros valores' -> 'Colar Valor(es) Antigos'
- Selecionar 'Continuar' / 'OK'



Recodificar valores extremos



Limpar a base de dados para o SPSS

Codificar Valores Ausentes



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

ANÁLISE DE DADOS EM GRH

Aula 5: As Mulheres Recebem o Mesmo que os Homens?

Fica para a próxima aula...