

# | Continuando a aula anterior

LIGUEM O SPSS E DESCARREGUEM OS  
MATERIAIS DAS AULAS 2 E 3

## Criação e Recodificação de Variáveis no SPSS

- ✓ *Adicionar Rótulos aos Valores da Variável*
- ✓ *Atribuir Valores a Variáveis Não Codificadas*
- x *Criar Variável Condicional*
- x *Criar Variável com Expressões Aritméticas*

## 4. Criação e Recodificação de Variáveis no SPSS

*Criar Variável Condicional*

## Recodificação de Variáveis

- A variável 'evaluation' mede a avaliação do trabalhador, pelo supervisor, numa escala de 0 a 100
- Objetivo:
  - Criar Variável Condicional que mede se a avaliação é superior a 50 ('positive\_eval')

0 - Não

1 - Sim

## Recodificação de Variáveis

- Selecionar 'Transformar' / 'Recodificar em Variáveis Diferentes'



\*ADGRH\_BD\_Aula4b\_final.sav [Conjunto de Dados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

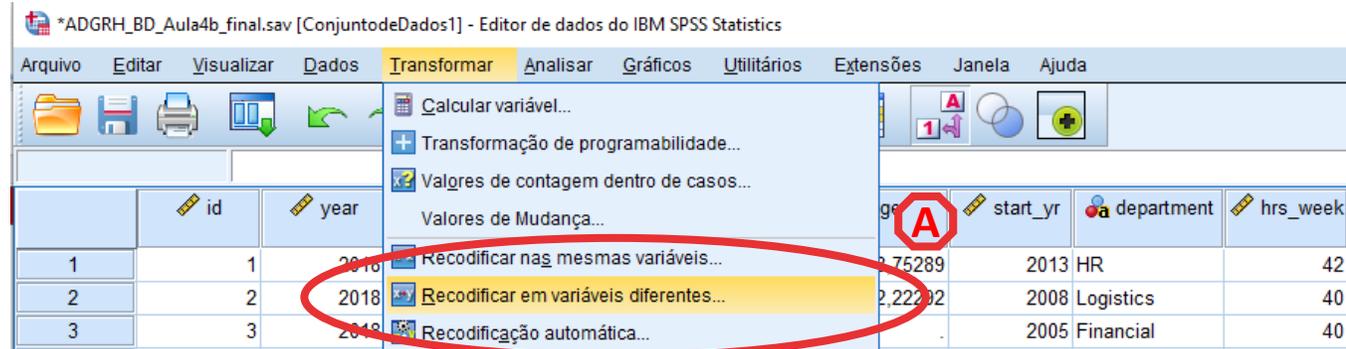
Calcular variável...  
Transformação de programabilidade...  
Valores de contagem dentro de casos...  
Valores de Mudança...  
Recodificar nas mesmas variáveis...  
**Recodificar em variáveis diferentes...**  
Recodificação automática...

	id	year	start_yr	department	hrs_week
1	1	2018	2013	HR	42
2	2	2018	2008	Logistics	40
3	3	2018	2005	Financial	40

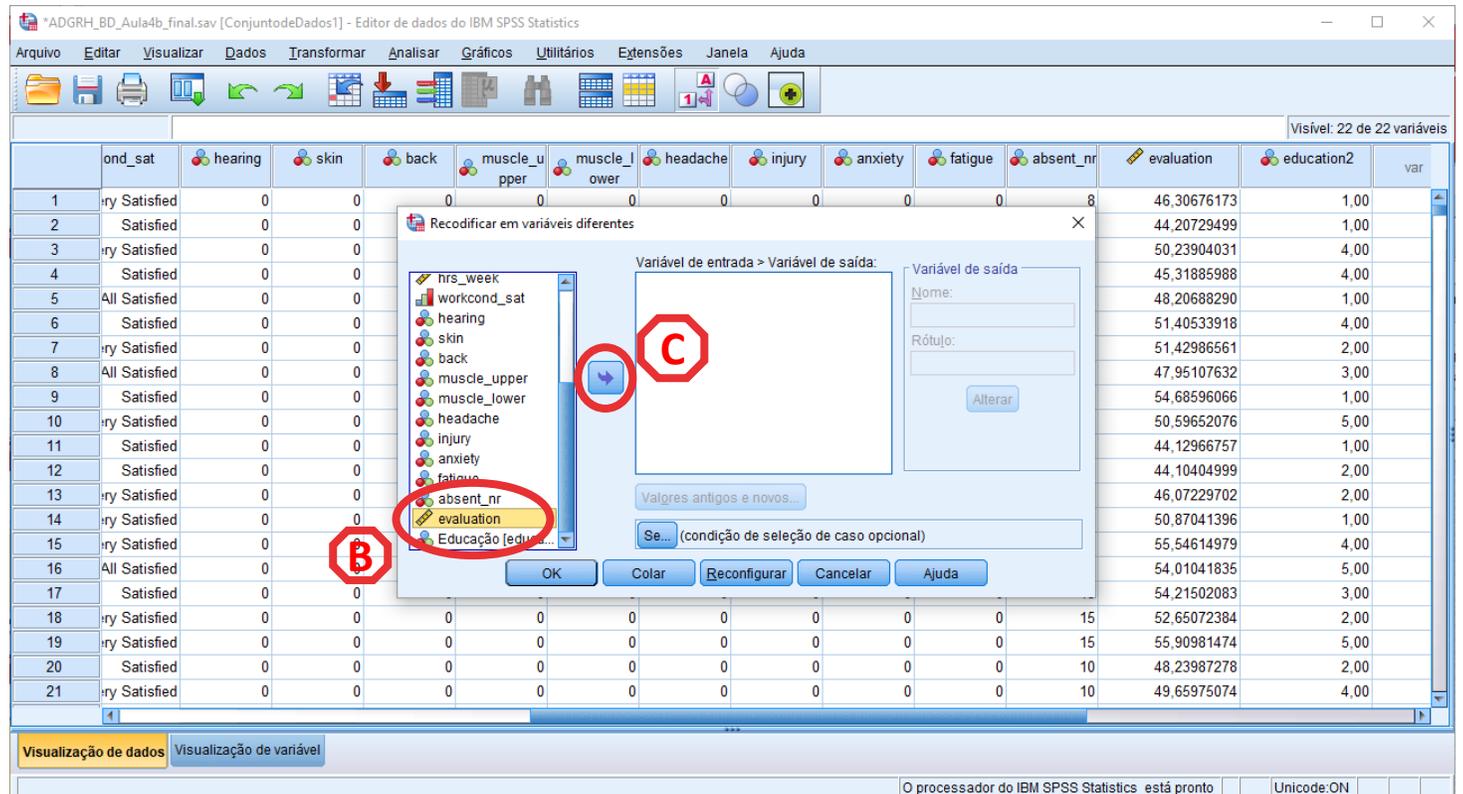
## Recodificação de Variáveis

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

A



B



C

## Recodificação de Variáveis

- Selecionar 'Transformar' / 'Recodificar em Variáveis Diferentes'
- Selecionar a variável 'evaluation'
- Definir 'evaluation' como Variável de Entrada
- Definir o nome da nova variável como 'positive\_eval' e o rótulo como 'Aval. Positiva'
- Selecionar 'Alterar'
- Selecionar 'Valores antigo e novo'

A

B

C

D

E

F

\*ADGRH\_BD\_Aula4b\_final.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.ON

## Recodificação de Variáveis

- Vamos primeiro definir que todos os casos que tem uma avaliação positiva (isto é,  $\geq 50$ ) na variável original **G**
- Associar-lhes o valor '1' (sim) **H** na nova variável
- Selecionar 'Incluir' **I**

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: Visível: 22 de 22 variáveis

Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo

Valor:

Omitido no sistema

Omitido para o sistema ou usuário

Amplitude:

até

Intervalo, MAIS BAIXO até o valor:

Intervalo, valor até o MAIS ALTO:

50

Todos os outros valores

Novo valor

Valor: 1

Omitido 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

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Recodificação de Variáveis

- Vamos primeiro definir que todos os casos que tem uma avaliação positiva (isto é,  $\geq 50$ ) na variável original **G**
- Associar-lhes o valor '1' (sim) **H** na nova variável
- Selecionar 'Incluir' **I**
- Selecionar 'Todos os outros valores' **J**
- Associar-lhes o valor '0' (não) **K** na nova variável
- Selecionar 'Incluir' **L**
- Selecionar 'Continuar' / 'OK' **M**

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: Visível: 22 de 22 variáveis

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, MAIS ALTO:

Todos os outros valores

Novo valor

Valor: **K**

Omisso no sistema

Copiar valor(es) antigo(s)

Antigo -> Novo:

50 thru Highest -> 1

Incluir **L**

Alterar

Remover

As variáveis de saída são sequências de caracteres Largura: 8

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

Continuar **M**  Cancelar  Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Recodificação de Variáveis

- Verificar se a variável está bem criada

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : positive\_eval ,0 Visível: 23 de 23 variáveis

	back	muscle_upper	muscle_lower	headache	injury	anxiety	fatigue	absent_nr	evaluation	education2	positive_eval	var	var
1	0	0	0	0	0	0	0	8	46,30676173	1,00	,00		
2	0	0	0	0	0	0	0	8	44,20729499	1,00	,00		
3	0	0	0	0	0	0	0	10	50,23904031	4,00	1,00		
4	0	0	0	0	0	0	0	9	45,31685988	4,00	,00		
5	0	0	0	0	0	0	0	9	48,20688290	1,00	,00		
6	0	0	0	0	0	0	0	8	51,40533918	4,00	1,00		
7	0	0	0	0	0	0	0	11	51,42986561	2,00	1,00		
8	0	0	0	0	0	1	0	6	47,95107632	3,00	,00		
9	0	0	0	0	0	0	0	14	54,68596066	1,00	1,00		
10	0	0	0	0	0	0	0	8	50,59652076	5,00	1,00		
11	0	0	0	0	0	0	0	10	44,12966757	1,00	,00		
12	0	0	0	0	0	0	0	8	44,10404999	2,00	,00		
13	0	0	0	0	0	0	0	11	46,07229702	2,00	,00		
14	0	0	1	0	0	0	0	9	50,87041396	1,00	1,00		
15	0	0	0	0	0	0	0	11	55,54614979	4,00	1,00		
16	0	0	0	0	0	1	0	13	54,01041835	5,00	1,00		
17	0	0	0	0	0	0	0	15	54,21502083	3,00	1,00		
18	0	0	0	0	0	0	0	15	52,65072384	2,00	1,00		
19	0	0	0	0	0	0	0	15	55,90981474	5,00	1,00		
20	0	0	0	0	0	0	0	10	48,23987278	2,00	,00		
21	0	0	0	0	0	0	0	10	49,65975074	4,00	,00		

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## 4. Criação e Recodificação de Variáveis no SPSS

*Criar Variável com Expressões Aritméticas (ex: subtração)*

## Recodificação de Variáveis

- **Objetivo:**
  - **Criar variável que mede o número de anos que a pessoa trabalha na empresa ('experience'):**

**'experience' = 'year' - 'start\_yr'**

## Recodificação de Variáveis

- Selecionar 'Transformar' / 'Calcular Variável'



\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Calcular variável...  
Transformação de programabilidade...  
Valores de contagem dentro de casos...  
Valores de Mudança...  
Recodificar nas mesmas variáveis...  
Recodificar em variáveis diferentes...  
Recodificação automática...

	id	year	age	start_yr	department	hrs_week
1	1	2018	3,75289	2013	HR	42
2	2	2018	2,22292	2008	Logistics	40
3	3	2018	.	2005	Financial	40

## Recodificação de Variáveis

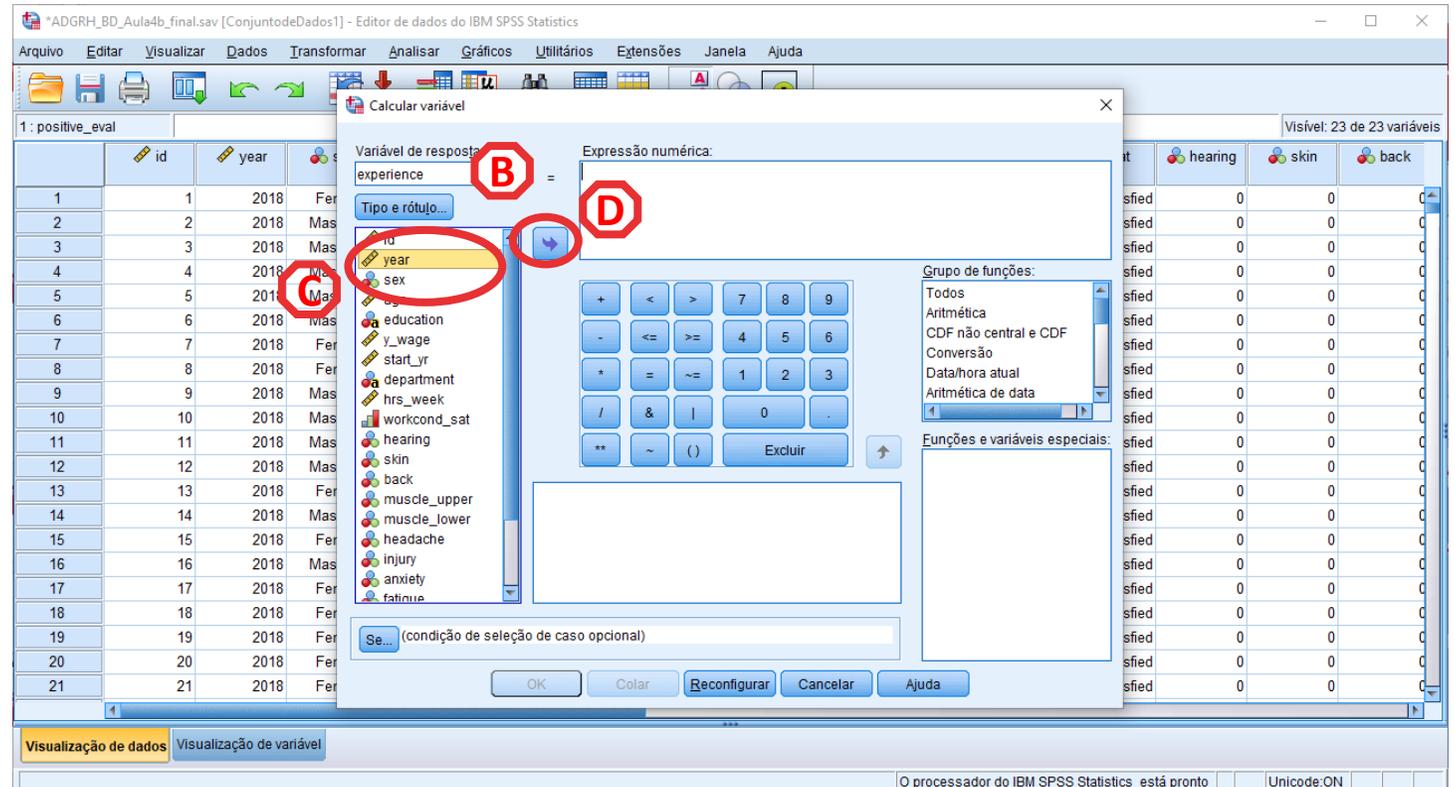
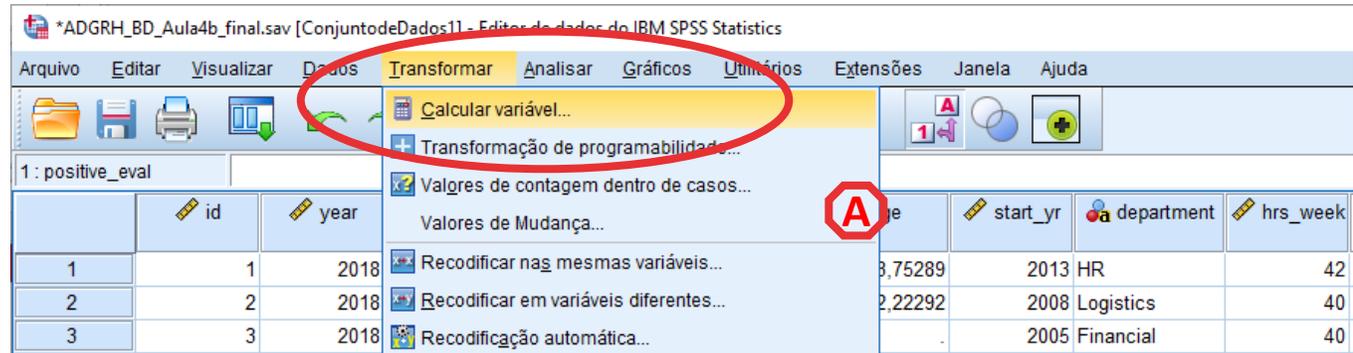
- Selecionar 'Transformar' / 'Calcular Variável'
- Nomear nova variável 'experience'
- Selecionar 'year'
- Enviar para a caixa

A

B

C

D



## Recodificação de Variáveis

- Selecionar 'Transformar' / 'Calcular Variável'
- Nomear nova variável 'experience'
- Selecionar 'year'
- Enviar para a caixa
- Selecionar operador '-'
- Selecionar 'start\_yr'
- Enviar para a caixa

A

B

C

D

E

F

G

The screenshot shows the IBM SPSS Statistics interface. The main window displays a data table with columns: id, year, sex, age, education, y\_wage, start\_yr, hrs\_week, workcond\_sat, hearing, skin, back, muscle\_upper, muscle\_lower, headache, injury, anxiety, and fatigue. The 'Calcular variável' dialog box is open, showing 'experience' as the 'Variável de resposta' and 'year' as the 'Expressão numérica'. The variable list on the left has 'year' and 'start\_yr' circled in red. The minus sign operator is also circled in red. The 'Calcular variável' dialog is overlaid on a data viewer showing a table with columns like 'id', 'year', 'sex', 'age', 'education', etc.

## Recodificação de Variáveis

- Seleccionar 'Transformar' / 'Calcular Variável'
- Nomear nova variável 'experience'
- Seleccionar 'year'
- Enviar para a caixa
- Seleccionar operador '-'
- Seleccionar 'start\_yr'
- Enviar para a caixa
- Seleccionar 'OK'

A

B

C

D

E

F

G

H

The screenshot shows the IBM SPSS Statistics interface. The main window is titled '\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics'. The 'Calcular variável' dialog box is open, showing the following details:

- Variável de resposta:** experience
- Expressão numérica:** year - start\_yr
- Grupo de funções:** Aritmética
- Funções e variáveis especiais:** (empty)

The 'OK' button is circled in red, indicating the final step in the process. The background data editor shows a table with columns: id, year, sex, age, education, y\_wage, start\_yr, department, hrs\_week, workcond\_sat, hearing, skin, back, muscle\_upper, muscle\_lower, headache, injury, anxiety, and faticioe.

## Recodificação de Variáveis

- A variável está bem criada?

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : experience 5,00 Visível: 24 de 24 variáveis

	headache	injury	anxiety	fatigue	absent_nr	evaluation	education2	positive_ev	experience	var	var	var	va
1	0	0	0	0	8	46,30676173	1,00	,00	5,00				
2	0	0	0	0	8	44,20729499	1,00	,00	10,00				
3	0	0	0	0	10	50,23904031	4,00	1,00	13,00				
4	0	0	0	0	9	45,31885988	4,00	,00	20,00				
5	0	0	0	0	9	48,20688290	1,00	,00	4,00				
6	0	0	0	0	8	51,40533918	4,00	1,00	6,00				
7	0	0	0	0	11	51,42986561	2,00	1,00	8,00				
8	0	0	1	0	6	47,95107632	3,00	,00	2,00				
9	0	0	0	0	14	54,68596066	1,00	1,00	6,00				
10	0	0	0	0	8	50,59652076	5,00	1,00	10,00				
11	0	0	0	0	10	44,12966757	1,00	,00	17,00				
12	0	0	0	0	8	44,10404999	2,00	,00	13,00				
13	0	0	0	0	11	46,07229702	2,00	,00	19,00				
14	0	0	0	0	9	50,87041396	1,00	1,00	1,00				
15	0	0	0	0	11	55,54614979	4,00	1,00	6,00				
16	0	0	1	0	13	54,01041835	5,00	1,00	18,00				
17	0	0	0	0	15	54,21502083	3,00	1,00	16,00				
18	0	0	0	0	15	52,65072384	2,00	1,00	16,00				
19	0	0	0	0	15	55,90981474	5,00	1,00	20,00				
20	0	0	0	0	10	48,23987278	2,00	,00	12,00				
21	0	0	0	0	10	49,65975074	4,00	,00	14,00				

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Recodificação de Variáveis

- A variável está bem criada?
- Para corrigir este erro, temos de ir para a área de 'Visualização de Variável'



\*ADGRH\_BD\_Aula4b\_final.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	Papel
4	age	Númerico	2	0		Nenhum	Nenhum	8	Direito	Escala	Entrada
5	education	Sequência ...	6	0		Nenhum	Nenhum	8	Esquerdo	Nominal	Entrada
6	y_wage	Númerico	11	5		Nenhum	Nenhum	13	Direito	Escala	Entrada
7	start_yr	Númerico	4	0		Nenhum	Nenhum	8	Direito	Escala	Entrada
8	department	Sequência ...	10	0		Nenhum	Nenhum	10	Esquerdo	Nominal	Entrada
9	hrs_week	Númerico	2	0		Nenhum	Nenhum	8	Direito	Escala	Entrada
10	workcond_sat	Númerico	8	2		{1,00, Not A...	Nenhum	15	Direito	Ordinal	Entrada
11	hearing	Númerico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
12	skin	Númerico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
13	back	Númerico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
14	muscle_upper	Númerico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
15	muscle_lower	Númerico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
16	headache	Númerico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
17	injury	Númerico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
18	anxiety	Númerico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
19	fatigue	Númerico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
20	absent_nr	Númerico	2	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
21	evaluation	Númerico	11	8		Nenhum	Nenhum	13	Direito	Escala	Entrada
22	education2	Númerico	8	2	Educação	Nenhum	Nenhum	12	Direito	Nominal	Entrada
23	positive_eval	Númerico	8	2	Aval. Positiva	Nenhum	Nenhum	15	Direito	Nominal	Entrada
24	experience	Númerico	8	2		Nenhum	Nenhum	12	Direito	Nominal	Entrada
25											
26											
..											

Visualização de dado Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON



## Recodificação de Variáveis

- A variável está bem criada?
- Para corrigir este erro, temos de ir para a área de 'Visualização de Variável'.
- Vamos mudar a Escala de Medida para 'Escala'



\*ADGRH\_BD\_Aula4b\_final.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	Papel
4	age	Numérico	2	0		Nenhum	Nenhum	8	Direito	Escala	Entrada
5	education	Sequência ...	6	0		Nenhum	Nenhum	8	Esquerdo	Nominal	Entrada
6	y_wage	Numérico	11	5		Nenhum	Nenhum	13	Direito	Escala	Entrada
7	start_yr	Numérico	4	0		Nenhum	Nenhum	8	Direito	Escala	Entrada
8	department	Sequência ...	10	0		Nenhum	Nenhum	10	Esquerdo	Nominal	Entrada
9	hrs_week	Numérico	2	0		Nenhum	Nenhum	8	Direito	Escala	Entrada
10	workcond_sat	Numérico	8	2		{1,00, Not A...	Nenhum	15	Direito	Ordinal	Entrada
11	hearing	Numérico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
12	skin	Numérico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
13	back	Numérico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
14	muscle_upper	Numérico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
15	muscle_lower	Numérico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
16	headache	Numérico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
17	injury	Numérico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
18	anxiety	Numérico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
19	fatigue	Numérico	1	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
20	absent_nr	Numérico	2	0		Nenhum	Nenhum	8	Direito	Nominal	Entrada
21	evaluation	Numérico	11	8		Nenhum	Nenhum	13	Direito	Escala	Entrada
22	education2	Numérico	8	2	Educação	Nenhum	Nenhum	12	Direito	Nominal	Entrada
23	positive_eval	Numérico	8	2	Aval. Positiva	Nenhum	Nenhum	15	Direito	Nominal	Entrada
24	experience	Numérico	8	2		Nenhum	Nenhum	12	Direito	Nominal	Entrada
25											
26											
..											

Visualização de dados Visualização de variável

## Recodificação de Variáveis

- A variável está bem criada?
- Para corrigir este erro, temos de ir para a área de 'Visualização de Variável'.
- Vamos mudar a Escala de Medida para 'Escala'
- Se voltarmos à área de 'Visualização de Dados'...

A

B

C

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Visual: 24 de 24 variáveis

	headache	injury	anxiety	fatigue	absent_nr	evaluation	education2	positive_eval	experience	var	var	var	va
1	0	0	0	0	8	46,30676173	1,00	,00	5,00				
2	0	0	0	0	8	44,20729499	1,00	,00	10,00				
3	0	0	0	0	10	50,23904031	4,00	1,00	13,00				
4	0	0	0	0	9	45,31885988	4,00	,00	20,00				
5	0	0	0	0	9	48,20688290	1,00	,00	4,00				
6	0	0	0	0	8	51,40533918	4,00	1,00	6,00				
7	0	0	0	0	11	51,42986561	2,00	1,00	8,00				
8	0	0	1	0	6	47,95107632	3,00	,00	2,00				
9	0	0	0	0	14	54,68596066	1,00	1,00	6,00				
10	0	0	0	0	8	50,59652076	5,00	1,00	10,00				
11	0	0	0	0	10	44,12966757	1,00	,00	17,00				
12	0	0	0	0	8	44,10404999	2,00	,00	13,00				
13	0	0	0	0	11	46,07229702	2,00	,00	19,00				
14	0	0	0	0	9	50,87041396	1,00	1,00	1,00				
15	0	0	0	0	11	55,54614979	4,00	1,00	6,00				
16	0	0	1	0	13	54,01041835	5,00	1,00	18,00				
17	0	0	0	0	15	54,21502083	3,00	1,00	16,00				
18	0	0	0	0	15	52,65072384	2,00	1,00	16,00				
19	0	0	0	0	15	55,90981474	5,00	1,00	20,00				
20	0	0	0	0	10	48,23987278	2,00	,00	12,00				
21	0	0	0	0	10	49,65975074	4,00	,00	14,00				

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

... verificamos que a escala da variável está correta

## 4. Criação e Recodificação de Variáveis no SPSS

*Criar Variável com Expressões Aritméricas (ex: média)*

## Recodificação de Variáveis

- **Objetivo:**
  - **Criar 'Índice de Problemas de Saúde' (Média)**
    - **Problemas de Audição ('hearing')**
    - **Problemas de Pele ('skin')**
    - **Dores nas Costas ('back')**
    - **Problemas Musculares – Parte Superior ('muscle\_upper')**
    - **Problemas Musculares – Parte Inferior ('muscle\_lower')**
    - **Dores de Cabeça ('headache')**

## Recodificação de Variáveis

- As variáveis do Índice são variáveis condicionais

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: hearing 0 Visível: 24 de 24 variáveis

	ge	education	y_wage	start_yr	department	hrs_week	workcond_sat	hearing	skin	back	muscle_upper	muscle_lower	headache
1	34	ISCED1	15883,75289	2013	HR	42	Not Very Satisfied	0	0	0	0	0	0
2	35	ISCED1	21082,22292	2008	Logistics	40	Satisfied	0	0	0	0	0	0
3	42	ISCED4	.	2005	Financial	40	Not Very Satisfied	0	0	0	0	0	0
4	42	ISCED4	28691,02479	1998	Production	42	Satisfied	0	0	0	0	0	0
5	35	ISCED1	21168,55747	2014	Production	41	Not At All Satisfied	0	0	0	0	0	0
6	38	ISCED4	26483,03628	2012	Sales	41	Satisfied	0	0	0	0	0	0
7	41	ISCED2	21787,73344	2010	Logistics	40	Not Very Satisfied	0	0	0	0	0	0
8	32	ISCED3	16982,64534	2016	Sales	39	Not At All Satisfied	0	0	0	0	0	0
9	43	ISCED1	26938,85363	2012	Production	40	Satisfied	0	0	0	0	0	0
10	40	ISCED5	28464,16163	2008	Marketing	39	Not Very Satisfied	0	0	0	0	0	0
11	40	ISCED1	24387,60008	2001	Production	40	Satisfied	0	0	0	0	0	0
12	37	ISCED2	23106,66895	2005	Production	40	Satisfied	0	0	0	0	0	0
13	43	ISCED2	23095,50876	1999	Production	38	Not Very Satisfied	0	0	0	0	0	0
14	36	ISCED1	21704,41879	2017	Logistics	39	Very Satisfied	0	0	0	1	0	0
15	42	ISCED4	25202,00497	2012	Logistics	39	Not Very Satisfied	0	0	0	0	0	0
16	49	ISCED5	34841,73692	2000	HR	39	Not At All Satisfied	0	0	0	0	0	0
17	48	ISCED3	28358,39067	2002	Sales	40	Satisfied	0	0	0	0	0	0
18	47	ISCED2	26537,40901	2002	Production	41	Not Very Satisfied	0	0	0	0	0	0
19	52	ISCED5	32857,75428	1998	Logistics	39	Not Very Satisfied	0	0	0	0	0	0
20	40	ISCED2	21341,50656	2006	Sales	41	Satisfied	0	0	0	0	0	0
21	42	ISCED4	24675,94203	2004	Financial	40	Very Satisfied	0	0	0	0	0	0

Visualização de dados Visualização de variável

O processador de dados do IBM SPSS Statistics está pronto. Unicode:ON

## Recodificação de Variáveis

- Selecionar 'Transformar' / 'Calcular Variável'

A

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Calcular variável...  
Transformação de programabilidade...  
Valores de contagem dentro de casos...  
Valores de Mudança...  
Recodificar nas mesmas variáveis...  
Recodificar em variáveis diferentes...  
Recodificação automática...

	id	year	age	start_yr	department	hrs_week
1	1	2018	3,75289	2013	HR	42
2	2	2018	2,22292	2008	Logistics	40
3	3	2018	.	2005	Financial	40

A

## Recodificação de Variáveis

- Selecionar 'Transformar' / 'Calcular Variável'
- Nomear nova variável 'health\_mean'

A

B

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Calcular variável

Variável de resposta: health\_mean

Expressão numérica:

Tipo e rótulo...

id year sex age education y\_wage start\_yr department hrs\_week workcond\_sat hearing skin back muscle\_upper muscle\_lower headache injury anxiety fatigue

Grupo de funções:  
Todos  
Aritmética  
CDF não central e CDF  
Conversão  
Data/hora atual  
Aritmética de data

Funções e variáveis especiais:  
Cvar  
Max  
Mean  
Median  
Min  
Sd  
Sum  
Variance

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

OK Colar Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

Visível: 24 de 24 variáveis

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Recodificação de Variáveis

- Selecionar 'Transformar' / 'Calcular Variável' **A**
- Nomear nova variável 'health\_mean' **B**
- Selecionar 'Grupo de Função' / 'Estatístico' **C**
- Selecionar função 'Mean' **D**
- Enviar para a caixa **E**



\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Calcular variável

Variável de resposta: health\_mean

Expressão numérica:

Tipo e rótulo...

id year sex age education y\_wage start\_lyr department hrs\_week workcond\_sat hearing skin back muscle\_upper muscle\_lower headache injury anxiety fatigue

Grupo de funções: Estatístico

Funções e variáveis especiais: Cvar Max Mean Median Min Sd Sum Variance

MEAN(numexpr,numexpr[...]). Numeric. Returns the arithmetic mean of its arguments that have valid, nonmissing values. This function requires two or more arguments, which must be numeric. You can specify a minimum number of valid arguments for this function to be evaluated.

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

## Recodificação de Variáveis

- Selecionar 'Transformar' / 'Calcular Variável'
- Nomear nova variável 'health\_mean'
- Selecionar 'Grupo de Função' / 'Estatístico'
- Selecionar função 'Mean'
- Enviar para a caixa

A

B

C

D

E

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Calcular variável

Variável de resposta: health\_mean

Expressão numérica: = MEAN(,?)

Tipo e rótulo...

id year sex age education y\_wage start\_yr department hrs\_week workcond\_sat hearing skin back muscle\_upper muscle\_lower headache injury anxiety fatigue

Grupo de funções:

- Significância
- Estatístico
- Escoragem
- Sequência de caracteres
- Criação da duração de tem
- Extração da duração de tem

Funções e variáveis especiais:

- CVar
- Max
- Mean
- Median
- Min
- Sd
- Sum
- Variance

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

## Recodificação de Variáveis

- Seleccionar 'Transformar' / 'Calcular Variável' **A**
- Nomear nova variável 'health\_mean' **B**
- Seleccionar 'Grupo de Função' / 'Estatístico' **C**
- Seleccionar função 'Mean' **D**
- Enviar para a caixa **E**
- Seleccionar a primeira variável do índice ('hearing') **F**
- Inserir a variável 'hearing' na função 'Mean' **G**



\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Calcular variável

Variável de resposta: health\_mean

Expressão numérica: MEAN(1?)

Tipo e rótulo...

id year sex age education y\_wage start\_yr department workweek workcond\_sat hearing skin back muscle\_upper muscle\_lower headache injury anxiety fatigue

Grupo de funções:

Significância  
Estatístico  
Escoragem  
Sequência de caracteres  
Criação da duração de tem  
Extração da duração de tem

Funções e variáveis especiais:

Cvar  
Max  
Mean  
Median  
Min  
Sd  
Sum  
Variance

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

## Recodificação de Variáveis

- **Exercício:** Inserir as restantes variáveis na função 'Mean'

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Calcular variável

Variável de resposta: health\_mean

Expressão numérica: MEAN(,?)

Tipo e rótulo...

id year sex age education y\_wage start\_yr department hrs\_week workcond\_sat skin back muscle\_upper muscle\_lower headache anxiety fatigue

MEAN(numexpr,numexpr[.]). Numeric. Returns the arithmetic mean of its arguments that have valid, nonmissing values. This function requires two or more arguments, which must be numeric. You can specify a minimum number of valid arguments for this function to be evaluated.

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

Grupo de funções: Significância Estatístico Escoragem Sequência de caracteres Criação da duração de tem Extração da duração de tem

Funções e variáveis especiais: CVar Max Mean Median Min Sd Sum Variance

OK Colar Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Recodificação de Variáveis

- **Exercício:** Inserir as restantes variáveis na função 'Mean'
- **Clicar OK**



\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Calcular variável

Variável de resposta: health\_mean

Expressão numérica: MEAN(hearing,skin,back,muscle\_upper,muscle\_lower,headache)

Grupo de funções:

- Significância
- Estatístico
- Escoragem
- Sequência de caracteres
- Criação da duração de tem
- Extração da duração de tem

Funções e variáveis especiais:

- Cvar
- Max
- Mean
- Median
- Min
- Sd
- Sum
- Variance

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

## Criação de variáveis: Lista de Funções

- A variável foi bem criada?

\*ADGRH\_BD\_Aula4b\_final.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : health\_mean ,0 Visível: 25 de 25 variáveis

	ety	fatigue	absent_nr	evaluation	education2	positive_eval	experience	health_mean	var	var	var	var	var
1	0	0	8	46,30676173	1,00	,00	5,00	,00					
2	0	0	8	44,20729499	1,00	,00	10,00	,00					
3	0	0	10	50,23904031	4,00	1,00	13,00	,00					
4	0	0	9	45,31885988	4,00	,00	20,00	,00					
5	0	0	9	48,20688290	1,00	,00	4,00	,00					
6	0	0	8	51,40533918	4,00	1,00	6,00	,00					
7	0	0	11	51,42986561	2,00	1,00	8,00	,00					
8	1	0	6	47,95107632	3,00	,00	2,00	,00					
9	0	0	14	54,68596066	1,00	1,00	6,00	,00					
10	0	0	8	50,59652076	5,00	1,00	10,00	,00					
11	0	0	10	44,12966757	1,00	,00	17,00	,00					
12	0	0	8	44,10404999	2,00	,00	13,00	,00					
13	0	0	11	46,07229702	2,00	,00	19,00	,00					
14	0	0	9	50,87041396	1,00	1,00	1,00	,17					
15	0	0	11	55,54614979	4,00	1,00	6,00	,00					
16	1	0	13	54,01041835	5,00	1,00	18,00	,00					
17	0	0	15	54,21502083	3,00	1,00	16,00	,00					
18	0	0	15	52,65072384	2,00	1,00	16,00	,00					
19	0	0	15	55,90981474	5,00	1,00	20,00	,00					
20	0	0	10	48,23987278	2,00	,00	12,00	,00					
21	0	0	10	49,65975074	4,00	,00	14,00	,00					

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Criação e Recodificação de Variáveis no SPSS

- ✓ *Adicionar Rótulos aos Valores da Variável*
- ✓ *Atribuir Valores a Variáveis Não Codificadas*
- ✓ *Criar Variável Condicional*
- ✓ *Criar Variável com Expressões Aritméticas*

## Exercícios em autonomia

- *Adicionar Rótulos aos Valores da Variável “fatigue”*
- *Atribuir Valores a categorias da variável “department”*
  1. Admin
  2. Audit
  3. Financial
  4. HR
  5. IT
  6. Logistics
  7. Marketing
  8. Production
  9. Sales
- *Criar Variável Condicional “cat\_eval” separando avaliações negativas (0-49), médias (50-70) e altas (70+)*
- *Criar Variável “index\_healthproblems” enquanto a soma das variáveis Expressões Aritméticas (hearing, skin, back, muscle\_upper, muscle\_lower, headache, injury, anxiety, fatigue)*

Aula 3:

## *'As Mulheres Recebem o Mesmo que os Homens?'*

Percentagens, médias, medianas e outras medidas de estatística descritiva

Docente: Daniela Craveiro  
dcraveiro@iseg.ulisboa.pt

- **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**

## 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**

	DEFINIÇÃO	TIPO DE VARIÁVEL	OBSERVAÇÕES
<b>Moda</b> (ou <i>valor modal</i> )	Valor mais frequente da variável	Todos	<ul style="list-style-type: none"> <li>• Adequada se a distribuição da variável é enviesada</li> <li>• Não é sensível a valores extremos</li> </ul>
<b>Média</b>	Valor médio das observações na variável	Razão	<ul style="list-style-type: none"> <li>• Adequada se distribuição da variável não é enviesada</li> <li>• É sensível a valores extremos</li> <li>• Quando os dados estão agrupados (ex: grupos de idade), calcula-se com base no valor mediano dos intervalos</li> </ul>
<b>Mediana</b>	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"> <li>• Adequada se a distribuição da variável é enviesada</li> <li>• Não é sensível a valores extremos</li> </ul>

## Medidas de Tendência Central

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

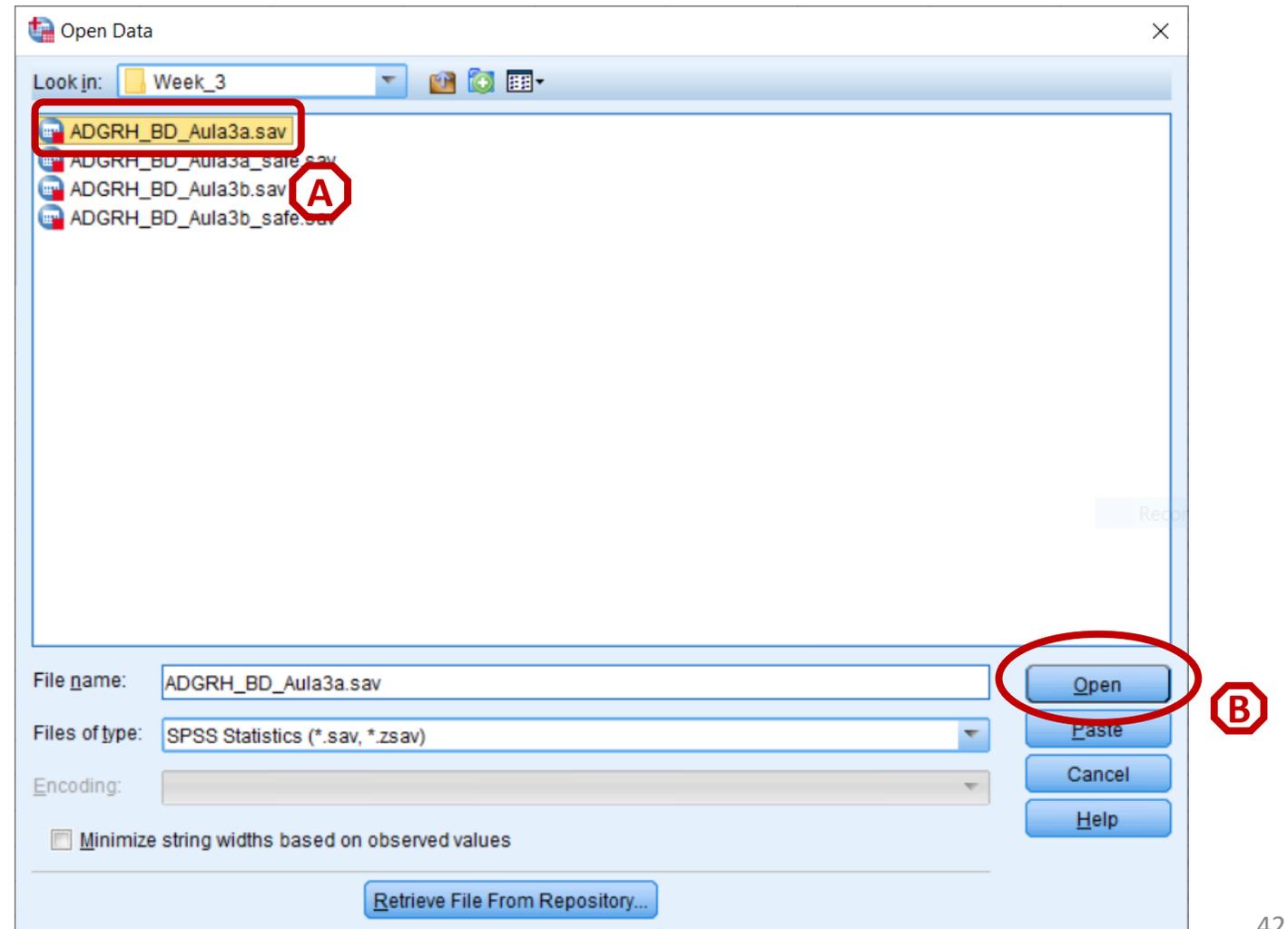
**Média**

**Mediana**

**Moda**

# Medidas de Tendência Central

- Selecionar ficheiro 'ADGRH\_BD\_Aula3a.sav' **A**
- Selecionar 'Abrir' **B**



# Medidas de Tendência Central

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

The image shows two screenshots of the IBM SPSS Statistics interface. The top screenshot shows the 'Analisar' menu with 'Estatística Descritiva' and 'Frequências...' highlighted. A red circle and the letter 'A' mark this menu path. The bottom screenshot shows the 'Frequências' dialog box with 'age' selected in the 'Variáveis:' list. Red circles and the letters 'B' and 'C' mark the selection of 'age' and the arrow button, respectively.

id	year	sex
1	2018	1
2	2018	2
3	2018	2
4	2018	2
5	2018	2

id	year	sex	age	education	y_wage	start_yr	department	hrs_week
1	2018	1	43	ISCED2	23095.50876	1999	Production	38
2	2018	2	36	ISCED1	21704.41879	2017	Logistics	39
3	2018	1	42	ISCED4	25202.00497	2012	Logistics	39

# Medidas de Tendência Central

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Frequências'



- Selecionar variável 'age'



- Colocar na caixa de 'Variáveis'



- Selecionar 'Estatísticas'



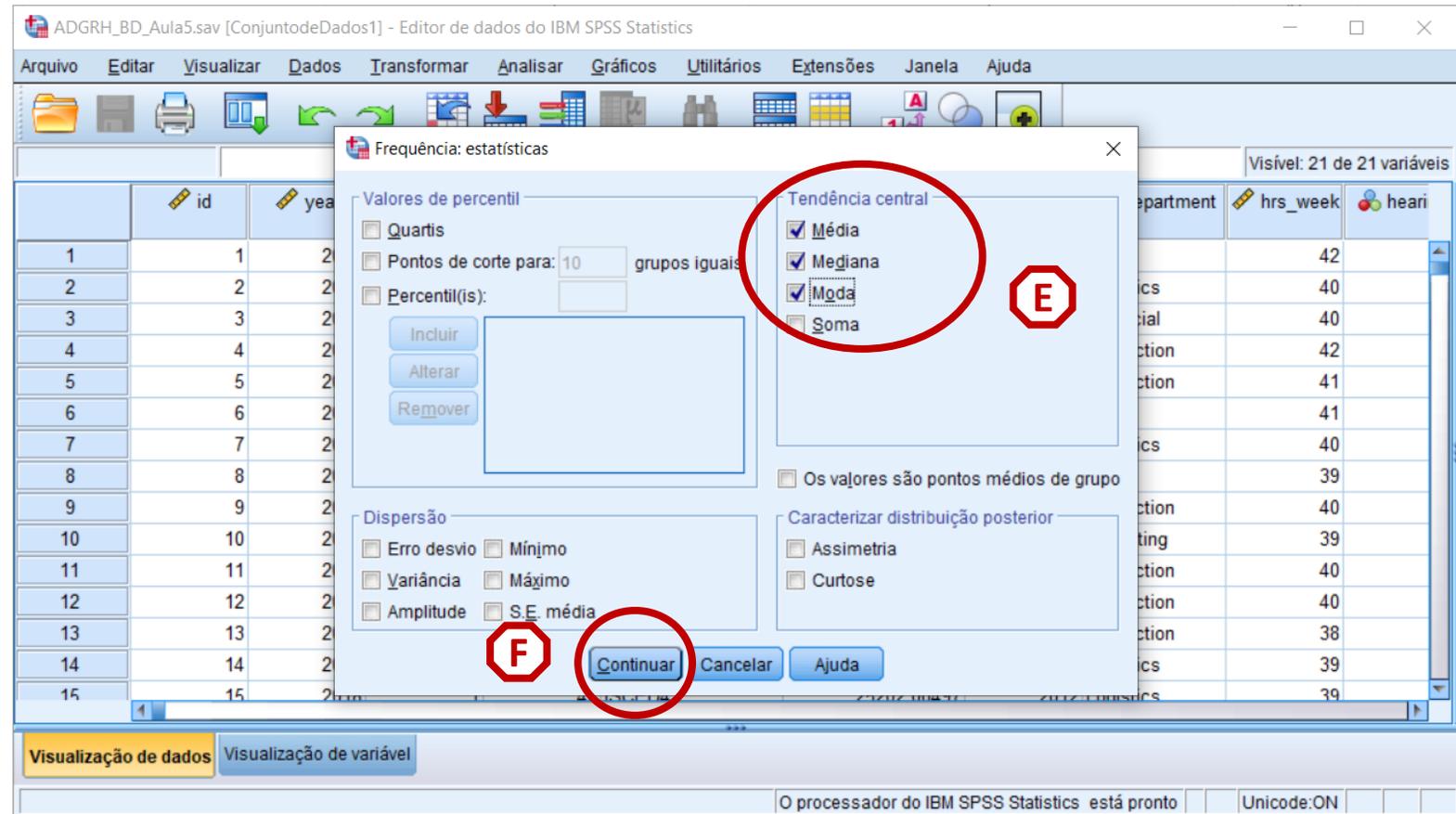
The screenshot shows the IBM SPSS Statistics interface. The top menu bar includes 'Arquivo', 'Editar', 'Visualizar', 'Dados', 'Transformar', 'Analisar', 'Gráficos', 'Utilitários', 'Extensões', 'Janela', and 'Ajuda'. The 'Analisar' menu is open, showing 'Estatística Descritiva' selected, with 'Frequências...' highlighted. A red circle highlights the 'Analisar' menu and its sub-menu. A red octagonal marker 'A' is placed to the right of the 'Frequências...' option.

The main data view shows a table with columns: id, year, sex, age, education, y\_wage, start\_yr, department, hrs\_week, and hearing. The 'Frequências' dialog box is open, showing 'age' in the 'Variável(is):' list. The 'Estatísticas...' button is highlighted with a red circle, and a red octagonal marker 'D' is placed to its right.

At the bottom of the dialog box, the 'Exibir tabelas de frequência' checkbox is checked. The 'OK' button is highlighted in yellow.

# 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 area displays the 'Frequências' (Frequencies) output for the variable 'age'. The statistics table shows a mean of 41.24, a median of 41.00, and a mode of 40. Below this, a detailed frequency table is provided, showing the distribution of ages from 21 to 36.

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

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

## Análise Exploratória dos Dados

### *Distribuição das Frequências*

- **Distribuição das Frequências**

	DEFINIÇÃO	TIPO DE VARIÁVEL	OBSERVAÇÕES
<b>Frequência</b>	Numero observações com um determinado valor na variável	Todos	-
<b>Proporção</b>	Número observações com um determinado valor na variável, dividido pelo número total de observações.	Todos	• Expressa numa escala de 0-1
<b>Percentagem</b>	Número observações com um determinado valor na variável, dividido pelo número total de observações, multiplicado por 100.	Todos	• Expressa numa escala de 0-100
<b>Rácio</b>	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)

## 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'

ADGRH\_BD\_Aula5.sav [ConjuntodeDados1] - 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

Frequências...  
Descritivos...  
Explorati...  
Tabela de referência cruzada...  
Razão...  
Gráficos P-P...  
Gráficos Q-Q...

id year sex

1	1	2018	1
2	2	2018	2
3	3	2018	2
4	4	2018	2
5	5	2018	2

ADGRH\_BD\_Aula5.sav [ConjuntodeDados1] - 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 heari

1 1 2018 20 2018 23095.50876 1999 Production 42

2 2 2018 20 2018 21704.41879 2017 Logistics 40

3 3 2018 20 2018 25202.00497 2012 Logistics 40

4 4 2018 20 2018 23095.50876 1999 Production 42

5 5 2018 20 2018 21704.41879 2017 Logistics 40

6 6 2018 20 2018 25202.00497 2012 Logistics 42

7 7 2018 20 2018 23095.50876 1999 Production 41

8 8 2018 20 2018 21704.41879 2017 Logistics 41

9 9 2018 20 2018 25202.00497 2012 Logistics 40

10 10 2018 20 2018 23095.50876 1999 Production 40

11 11 2018 20 2018 21704.41879 2017 Logistics 39

12 12 2018 20 2018 25202.00497 2012 Logistics 40

13 13 2018 20 2018 23095.50876 1999 Production 38

14 14 2018 20 2018 21704.41879 2017 Logistics 39

15 15 2018 20 2018 25202.00497 2012 Logistics 39

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

# Distribuição de Frequências

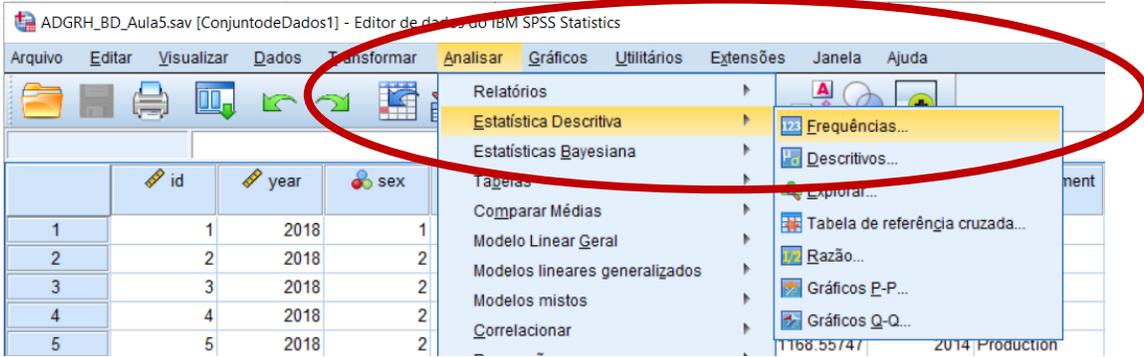
- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Frequências'
- Selecionar variável 'department'
- Colocar na caixa de 'Variáveis'
- Selecionar 'OK'

A

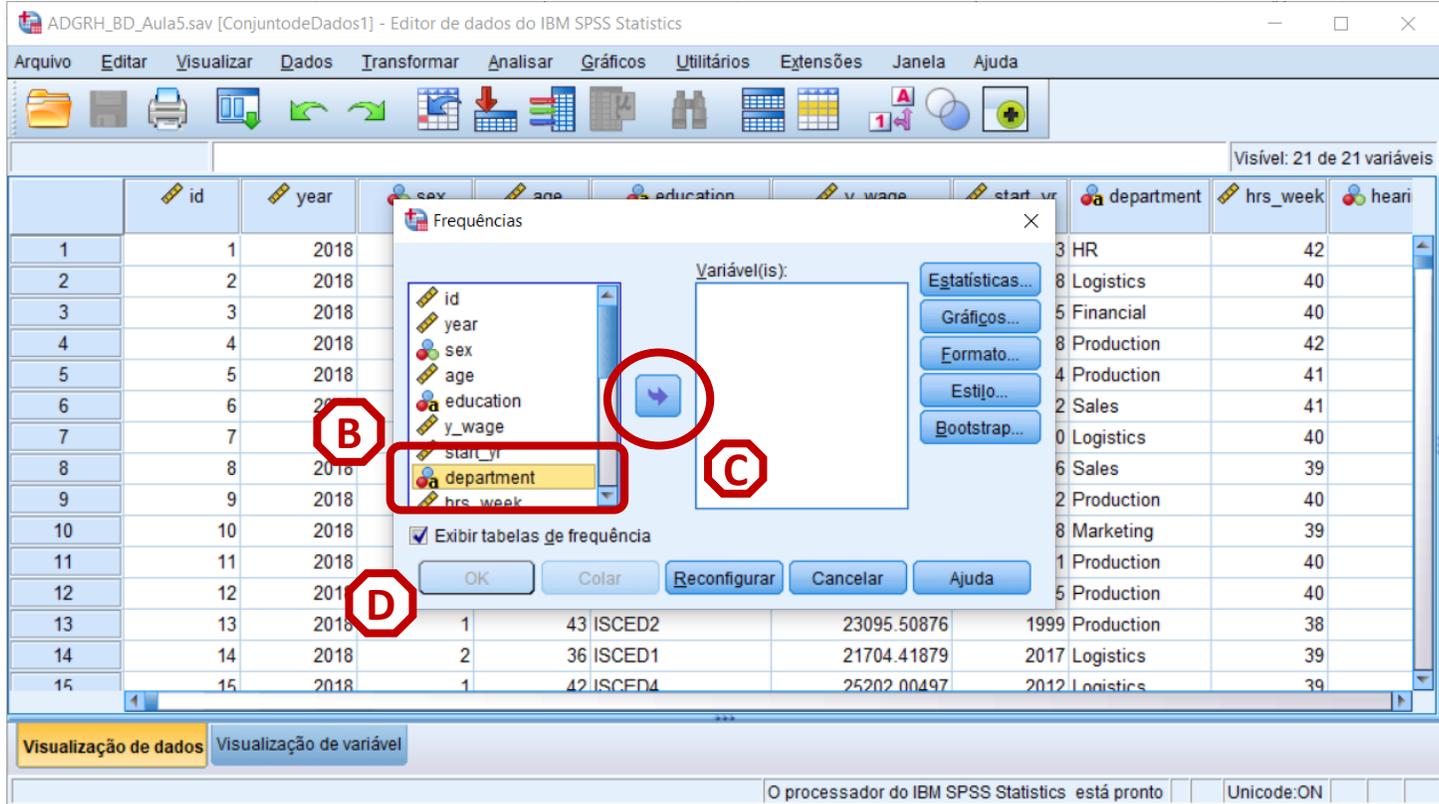
B

C

D



A



# Tabela de Frequências

- O resultado é publicado no 'Visualizador de Resultados'

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

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

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

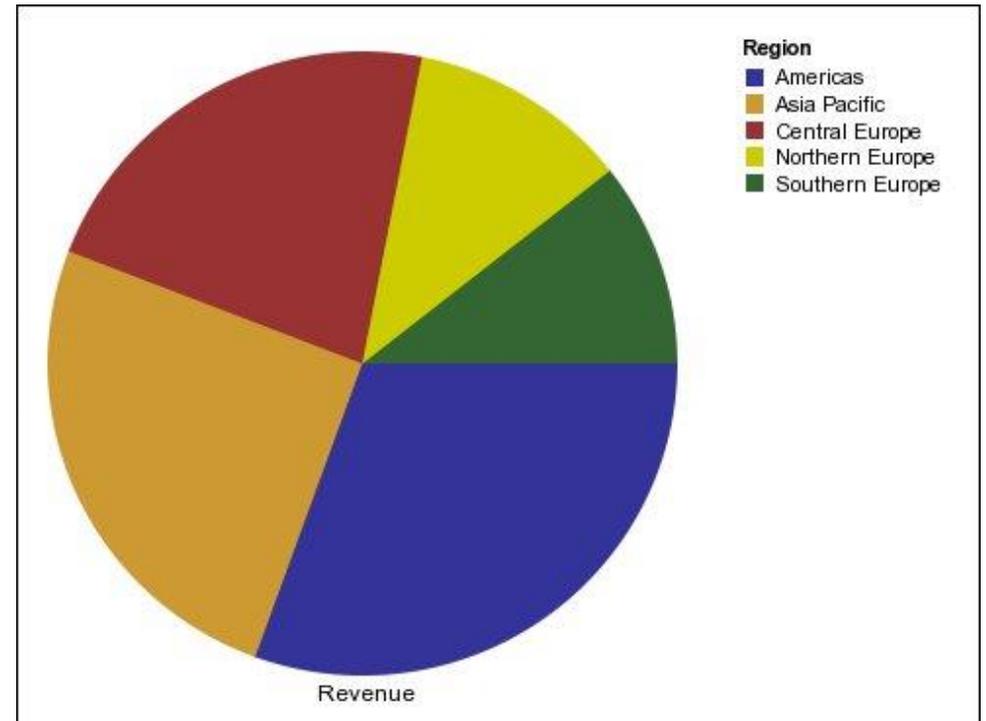
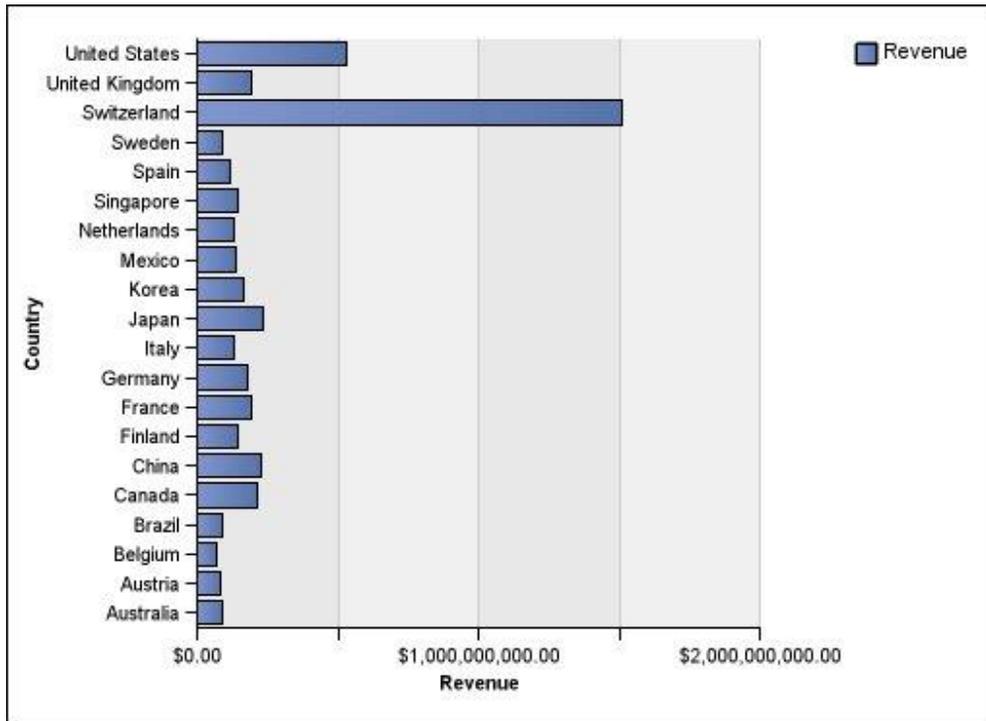
Unicode:ON

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

Variável  
Nominal  
ou Ordinal

Gráfico de Barras

Gráfico Circular

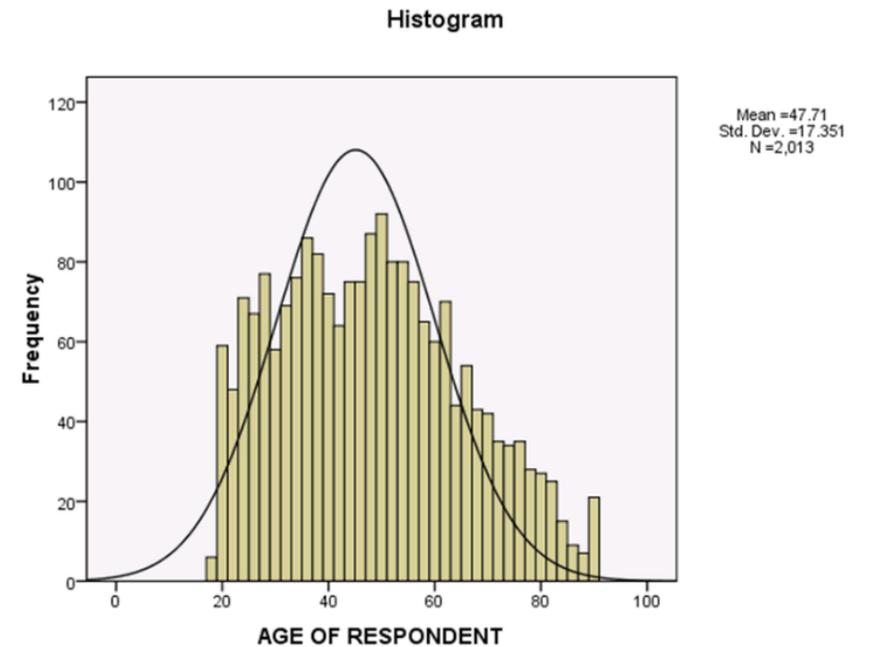
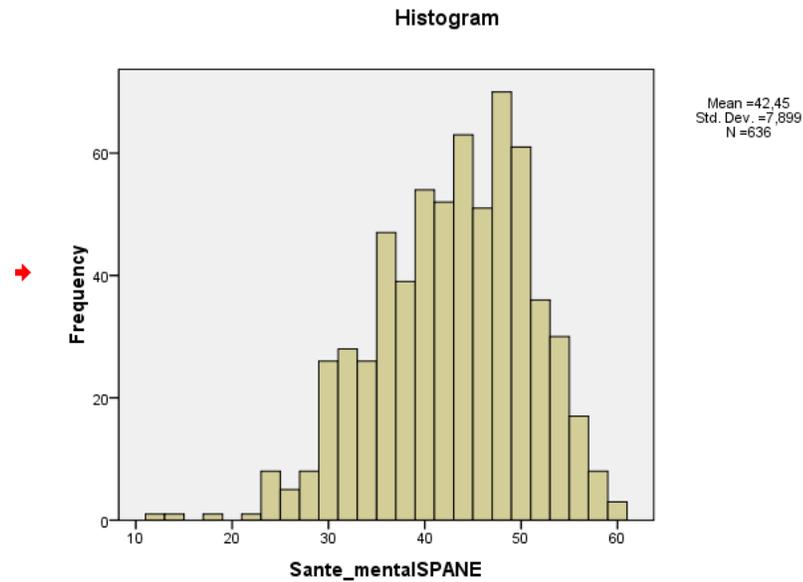


- Qual a melhor forma de visualizar a distribuição de frequências?

Variável  
Intervalar  
ou Razão

Histograma

Histograma,  
com Curva de Distribuição



# Gráfico de Barras

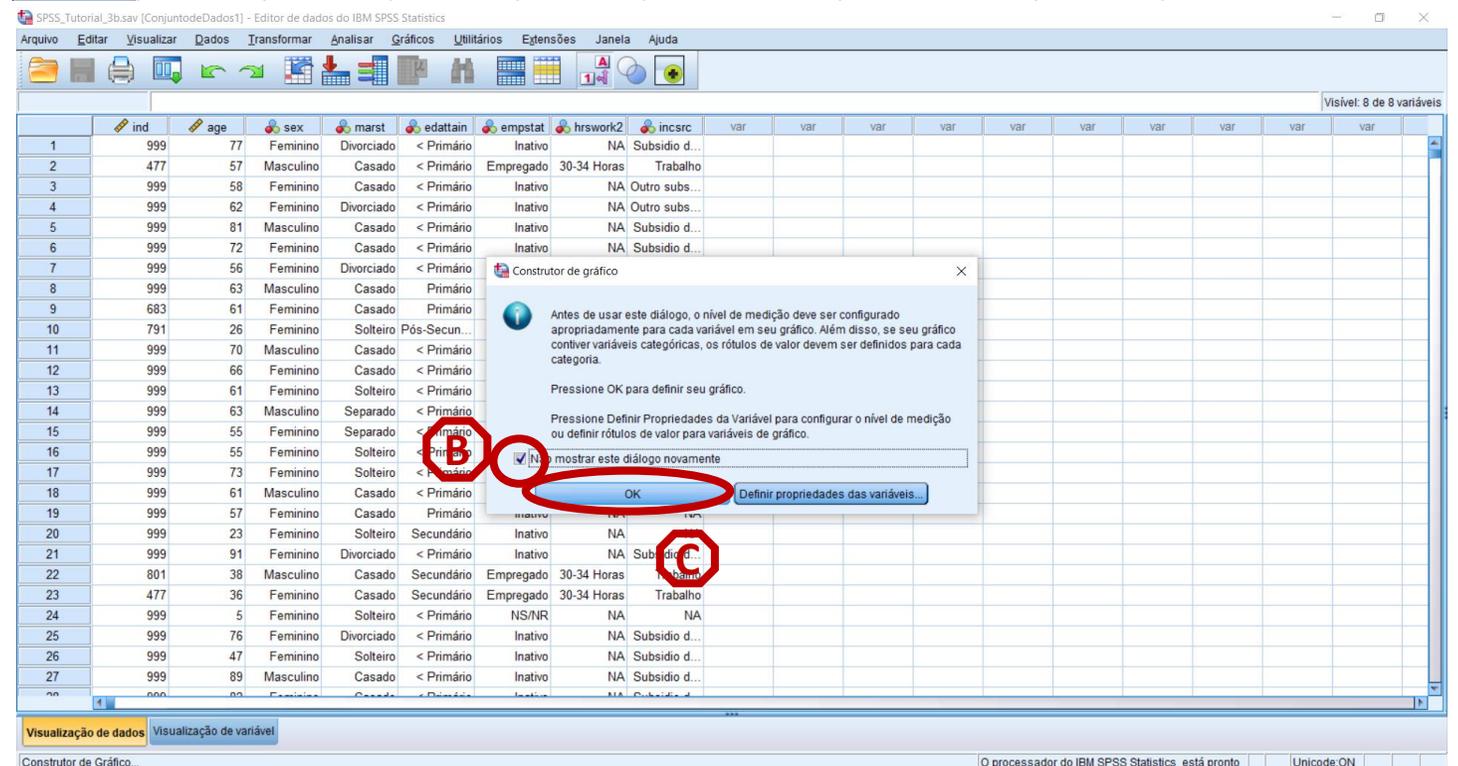
- Selecionar 'Gráficos' / 'Construtor de Gráfico'



- Selecionar 'Não mostrar este diálogo novamente'

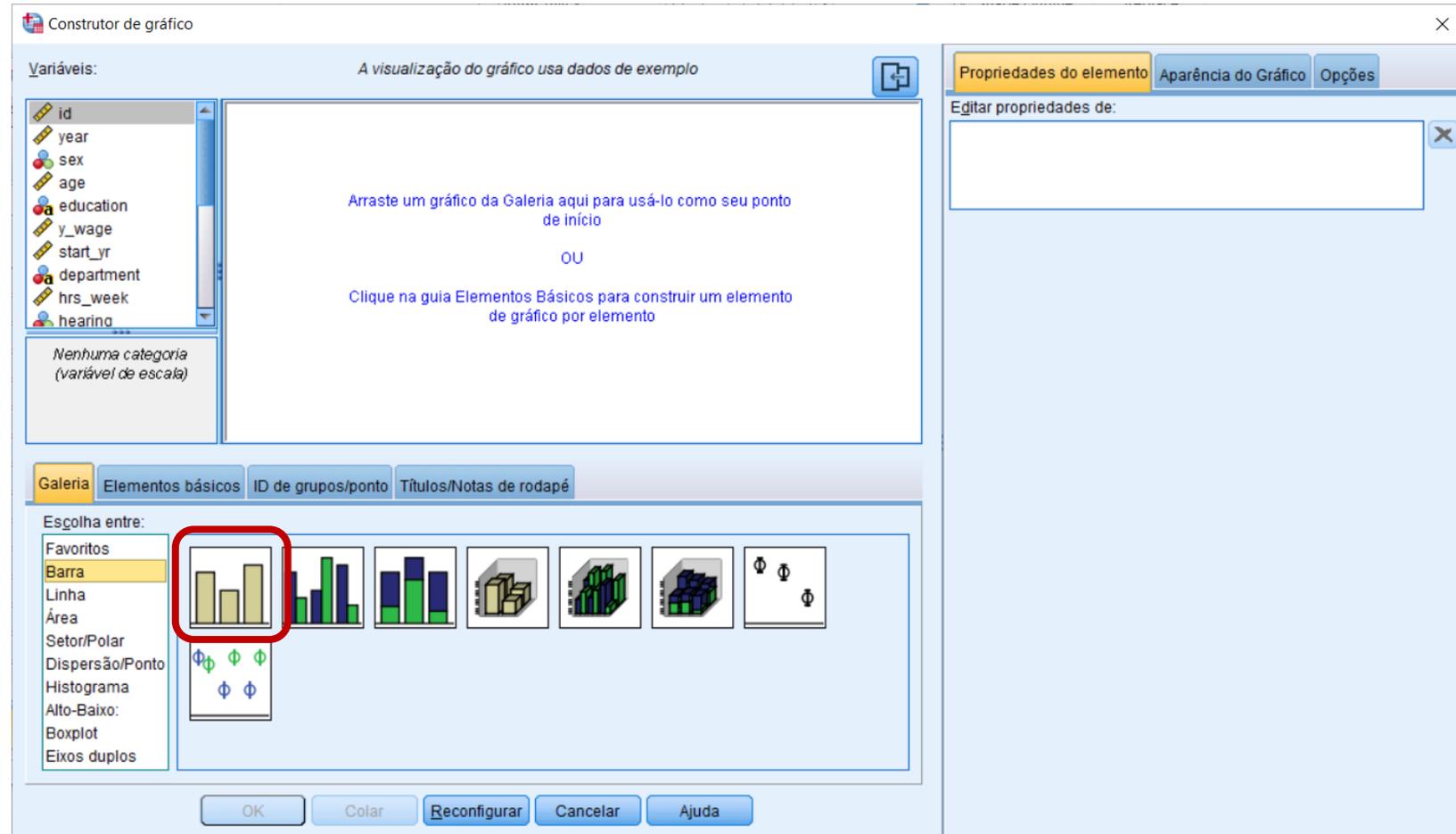


- Selecionar 'OK'



# 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

Barra Simples Contagem

Eixo Y?

Filtro?

Eixo X?

Galeria | Elementos básicos | ID de grupos/pontos | Legendas/Notas de rodapé

Escolha entre:

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

Propriedades do elemento | Aparência do Gráfico | Opções

Editar propriedades de:

Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

Título 1

Estadísticas

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

OK | Colar | Reconfigurar | Cancelar | Ajuda

# 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

The screenshot shows the 'Construtor de gráfico' (Chart Builder) window. The main area displays a simple bar chart titled 'Barra Simples Contagem de department'. The y-axis is labeled 'Contagem' and the x-axis is labeled 'department'. The chart shows bars for Production, Logistics, Sales, Admin, Financial, Marketing, IT, HR, and Audit. The 'Admin' department has the highest count.

On the left, the 'Variáveis:' list includes: year, sex, age, education, start\_yr, start\_age, experience, y\_wage, department, and hrs\_week. The 'department' variable is selected.

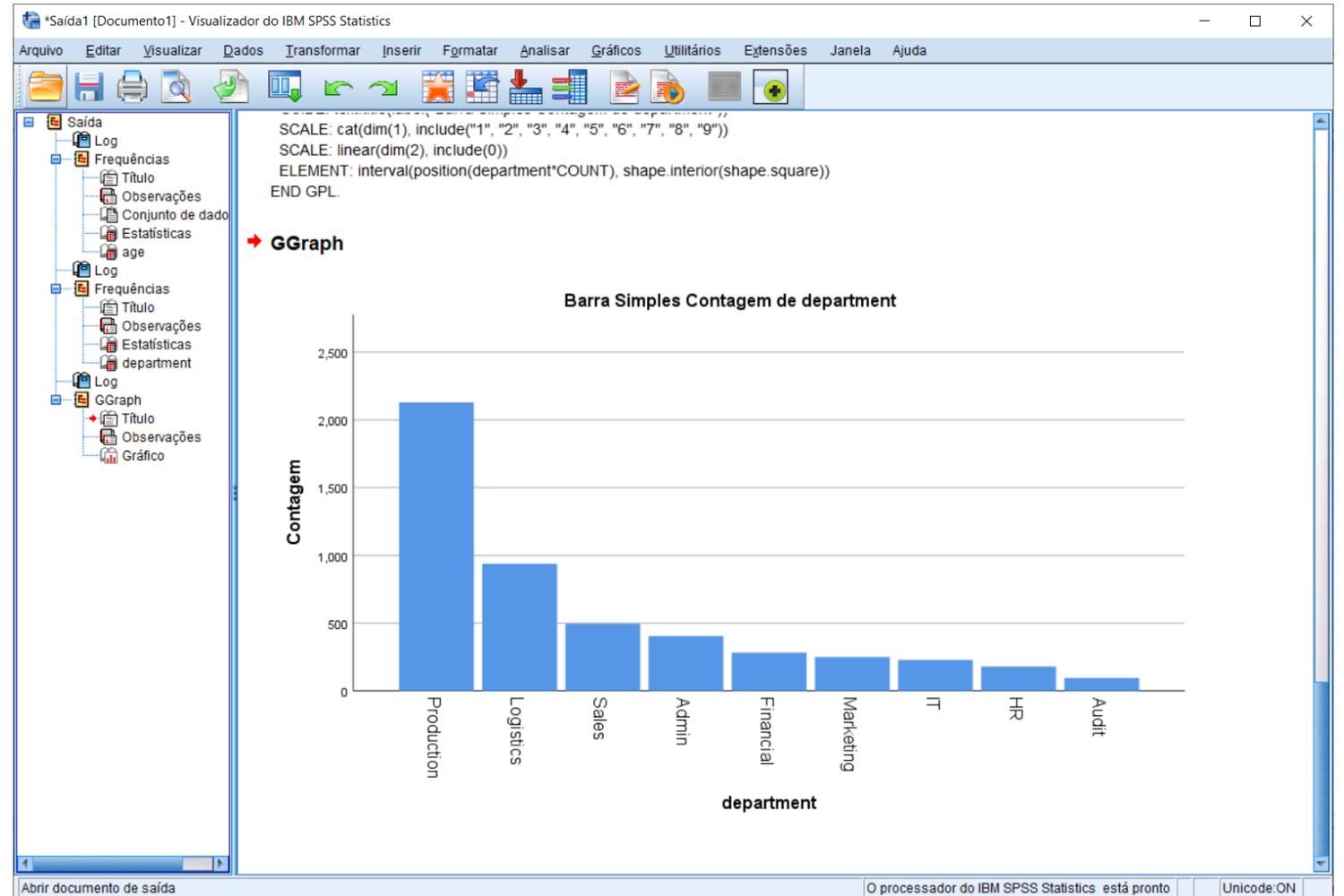
Below the chart, the 'Galeria' (Gallery) tab is active, showing various chart types. The 'Barra' (Bar) chart type is selected.

On the right, the 'Propriedades do elemento' (Element Properties) panel is open, showing the 'Barra1' element. The 'Estatísticas' (Statistics) section is visible, with 'Variável:' set to 'Contagem' and 'Estatística:' set to 'Contagem'. The 'Exibir barra de erros' (Show error bars) checkbox is checked.

At the bottom, the 'OK' button is circled in red.

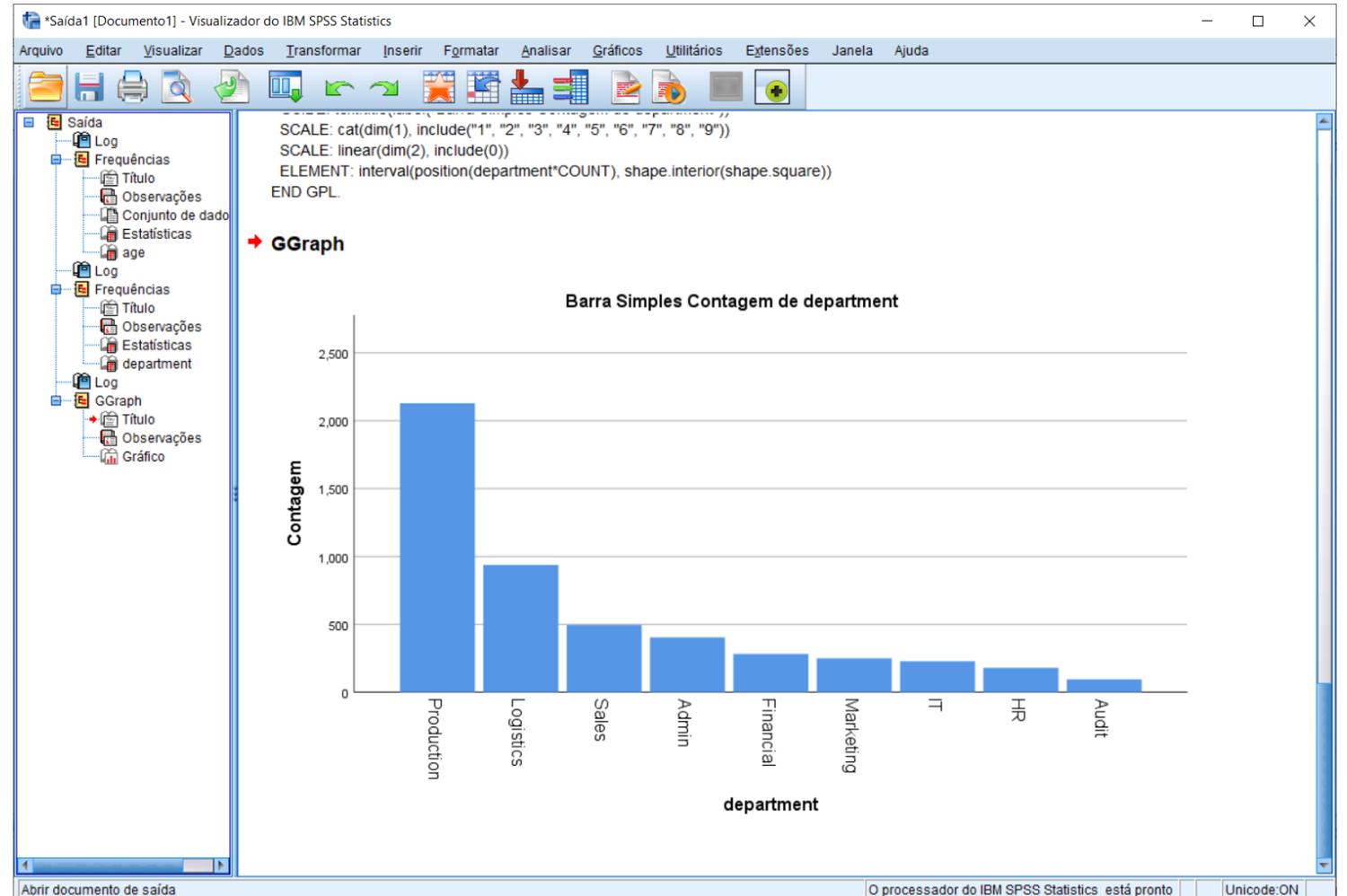
# Gráfico de Barras

- O gráfico é publicado no 'Visualizador de Resultados'



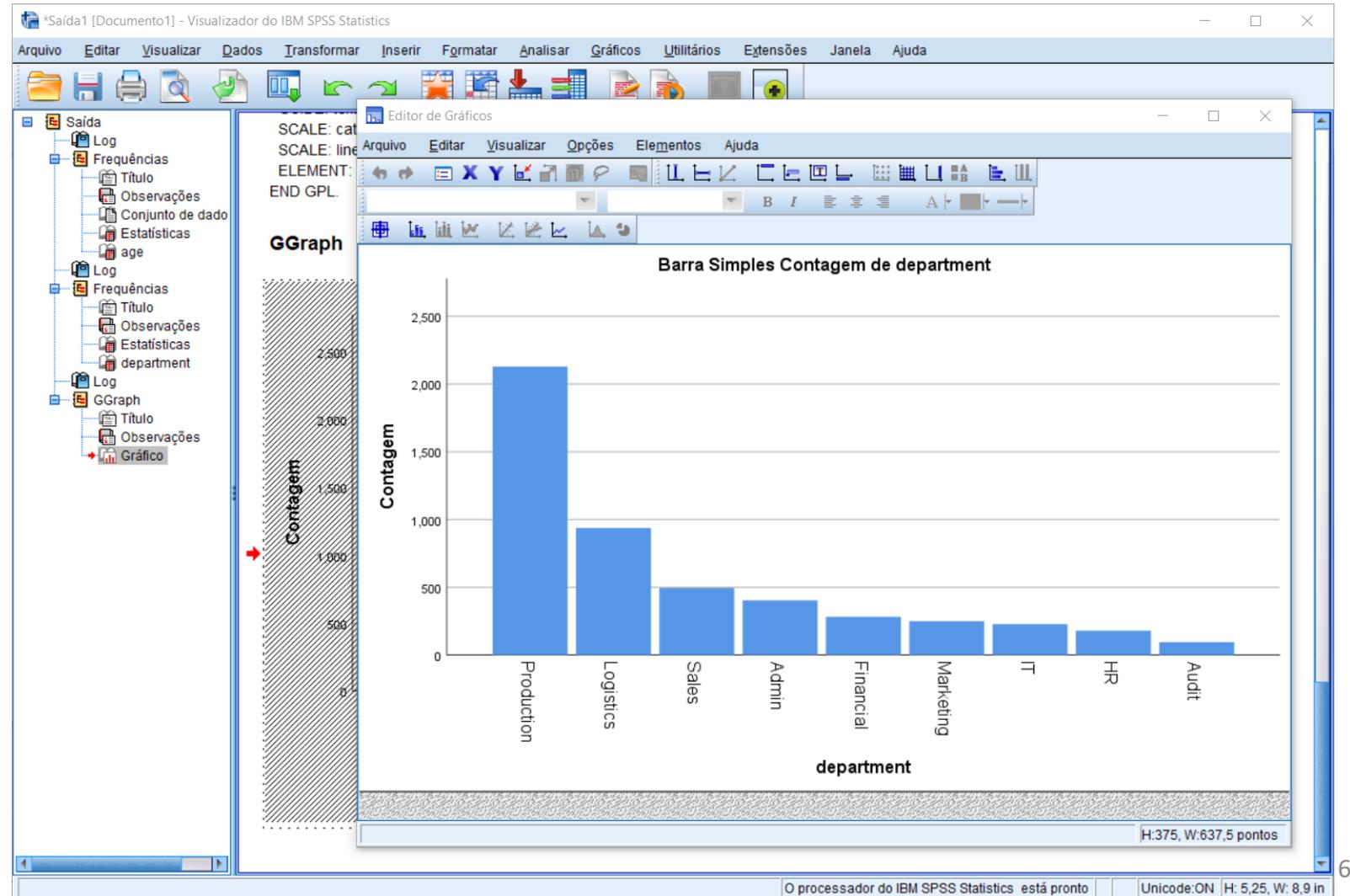
# Gráfico de Barras

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



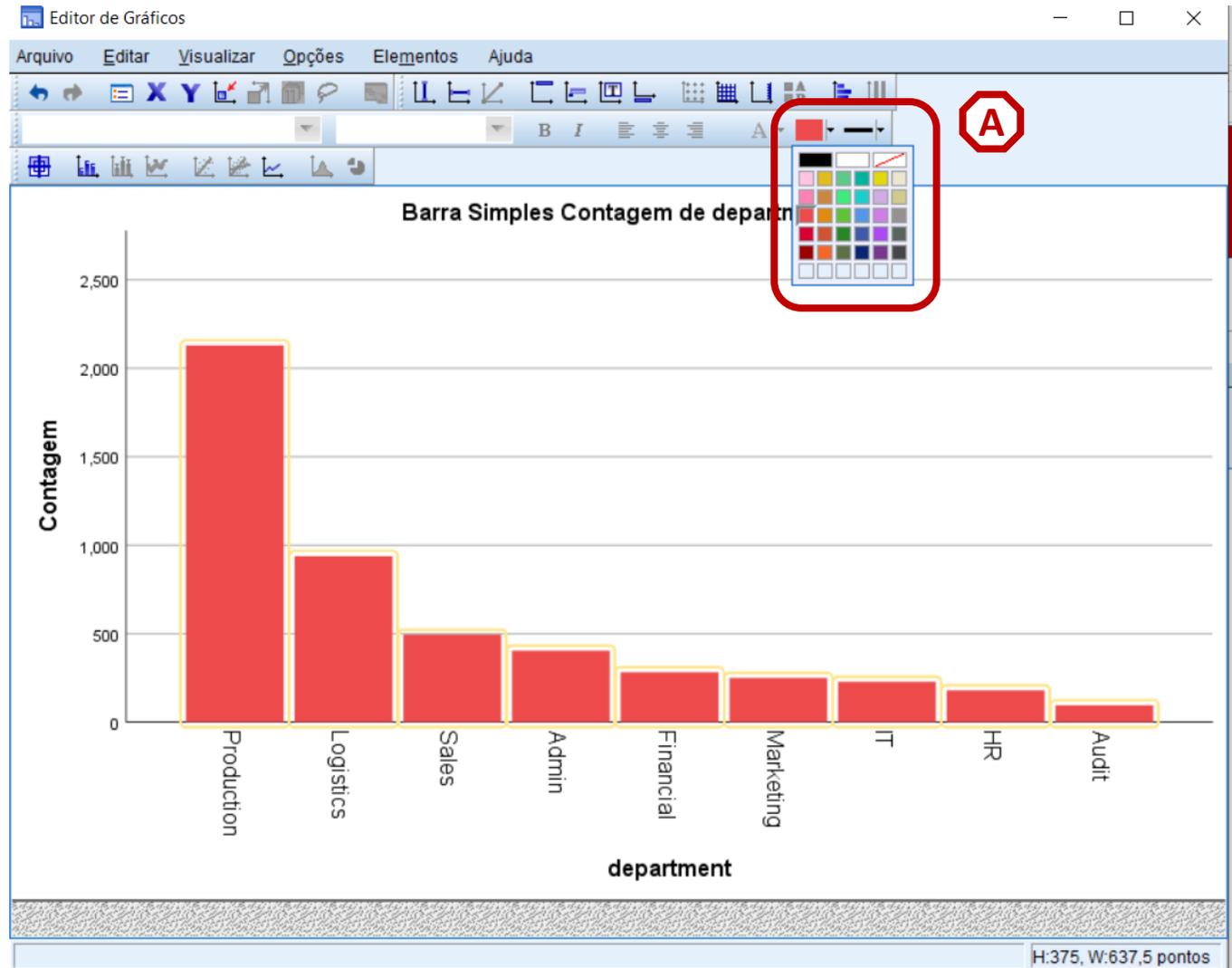
# Gráfico de Barras

- 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*

- **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
<b>Intervalo de Variação</b>	Diferença entre o valor mais alto e o valor mais baixo da variável	Intervalar* Razão	<ul style="list-style-type: none"><li>• É sensível a valores extremos</li></ul>

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

## • Medidas de dispersão

	DEFINIÇÃO	TIPO DE VARIÁVEL	OBSERVAÇÕES
<b>Intervalo de Variação</b>	Diferença entre o valor mais alto e o valor mais baixo da variável	Intervalar* Razão	<ul style="list-style-type: none"> <li>• É sensível a valores extremos</li> </ul>
<b>Amplitude Interquartis</b>	Diferença entre o valor do 3º Quartil e o valor do 1º Quartil	Intervalar* Razão	<ul style="list-style-type: none"> <li>• Não é sensível a valores extremos</li> </ul>
<b>Rácio Interdecis</b>	90/10: Valor do 9º decil dividido pelo valor do 1º decil	Intervalar* Razão	<ul style="list-style-type: none"> <li>• Alternativa: 80/20</li> <li>• Não é sensível a valores extremos</li> </ul>
<b>Variância</b>	Média das diferenças dos valores observados em relação à média, elevadas ao quadrado	Intervalar* Razão	<ul style="list-style-type: none"> <li>• = 0: Não há variância</li> <li>• Quanto maior, maior variância</li> <li>• É sensível a valores extremos</li> <li>• Difícil de interpretar</li> </ul>
<b>Desvio-Padrão</b>	Raiz quadrada da variância.	Intervalar* Razão	<ul style="list-style-type: none"> <li>• Mais fácil de interpretar</li> <li>• Usa a mesma escala de valores da variável</li> <li>• É sensível a valores extremos</li> </ul>

## Medidas de Dispersão

- **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'

A

B

C

D

ADGRH\_BD\_Aula4b.sav [ConjuntodeDados1] - 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

Frequências...  
Descritivos...  
Explorar...  
Tabela de referência cruzada...  
Razão...  
Gráficos P-P...  
Gráficos Q-Q...

ADGRH\_BD\_Aula4b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

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

muscle\_upper  
muscle\_lower  
headache  
injury  
anxiety  
fatigue  
absent\_nr  
workcond\_sat  
**evaluation**

Variável(is):

Estatísticas...  
Gráficos...  
Formato...  
Estilo...  
Bootstrap...

Exibir tabelas de frequência

OK Color Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

# 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

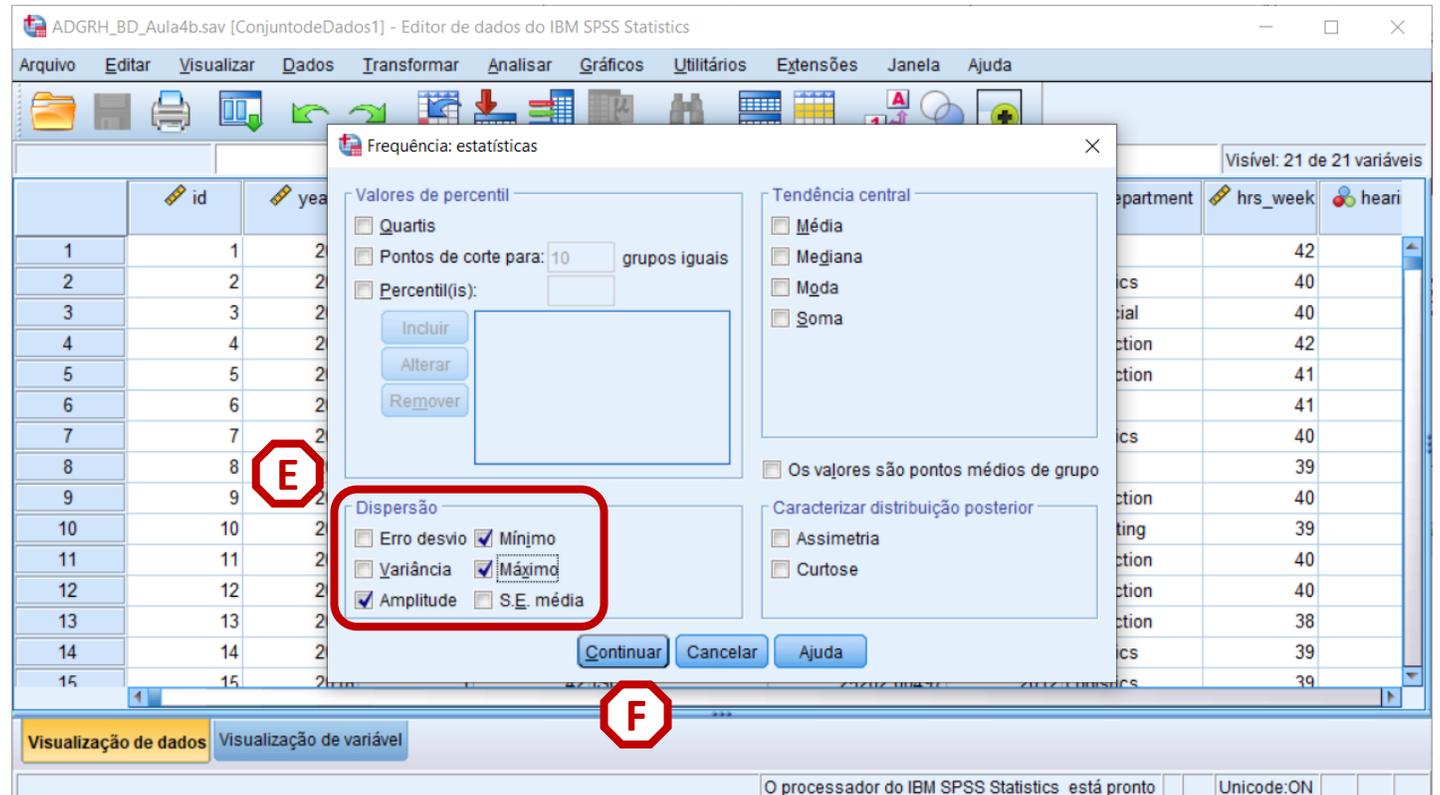
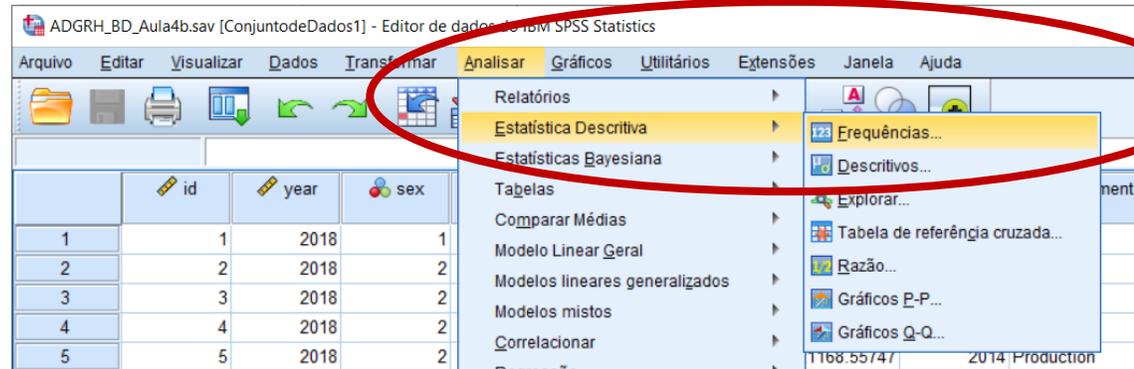
B

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 'Visualizador do IBM SPSS Statistics' window. The main content area displays the following command and results:

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

**→ Frequências**

**Estadísticas**

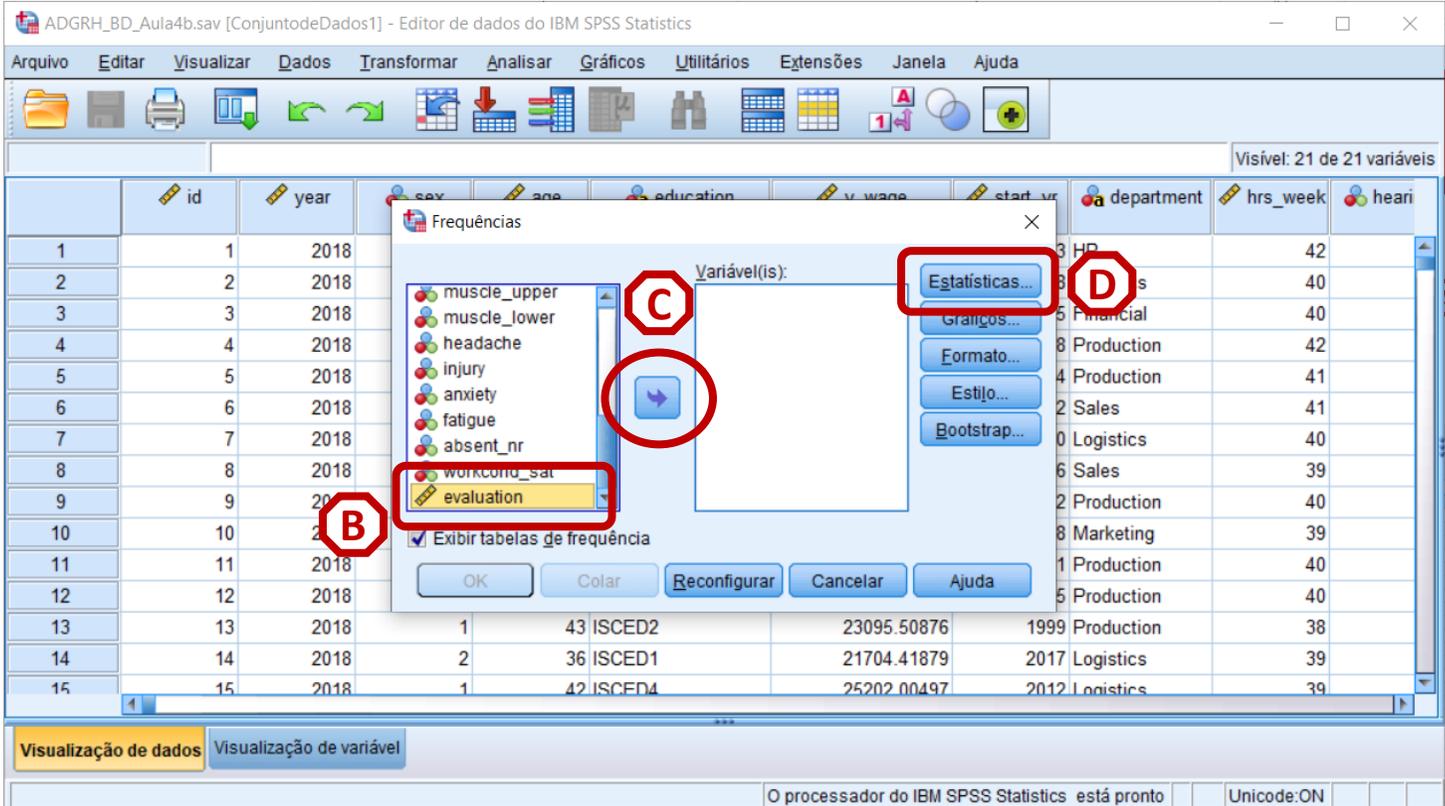
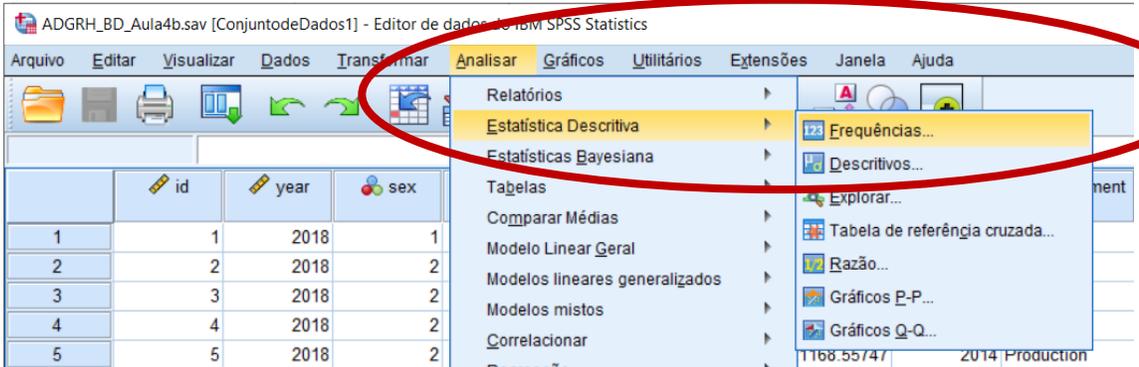
evaluation		
N	Válido	5000
	Omisso	0
Intervalo		19.73
Mínimo		41.85
Máximo		61.58

**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'



# 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

The screenshot shows the IBM SPSS Statistics interface with the 'Frequência: estatísticas' dialog box open. The 'Valores de percentil' section has the 'Quartis' checkbox checked. The 'Tendência central' section has 'Média', 'Mediana', 'Moda', and 'Soma' unchecked. The 'Dispersão' section has 'Erro desvio', 'Mínimo', 'Variância', 'Máximo', 'Amplitude', and 'S.E. média' unchecked. The 'Caracterizar distribuição posterior' section has 'Assimetria' and 'Curtose' unchecked. The 'Os valores são pontos médios de grupo' checkbox is also unchecked. The 'Continuar', 'Cancelar', and 'Ajuda' buttons are at the bottom of the dialog box. A red arrow points from the 'E' label to the 'Quartis' checkbox.

# Medidas de Dispersão

- Selecionar 'Percentis'



- Inserir 10º Percentil (= 1º Decil)

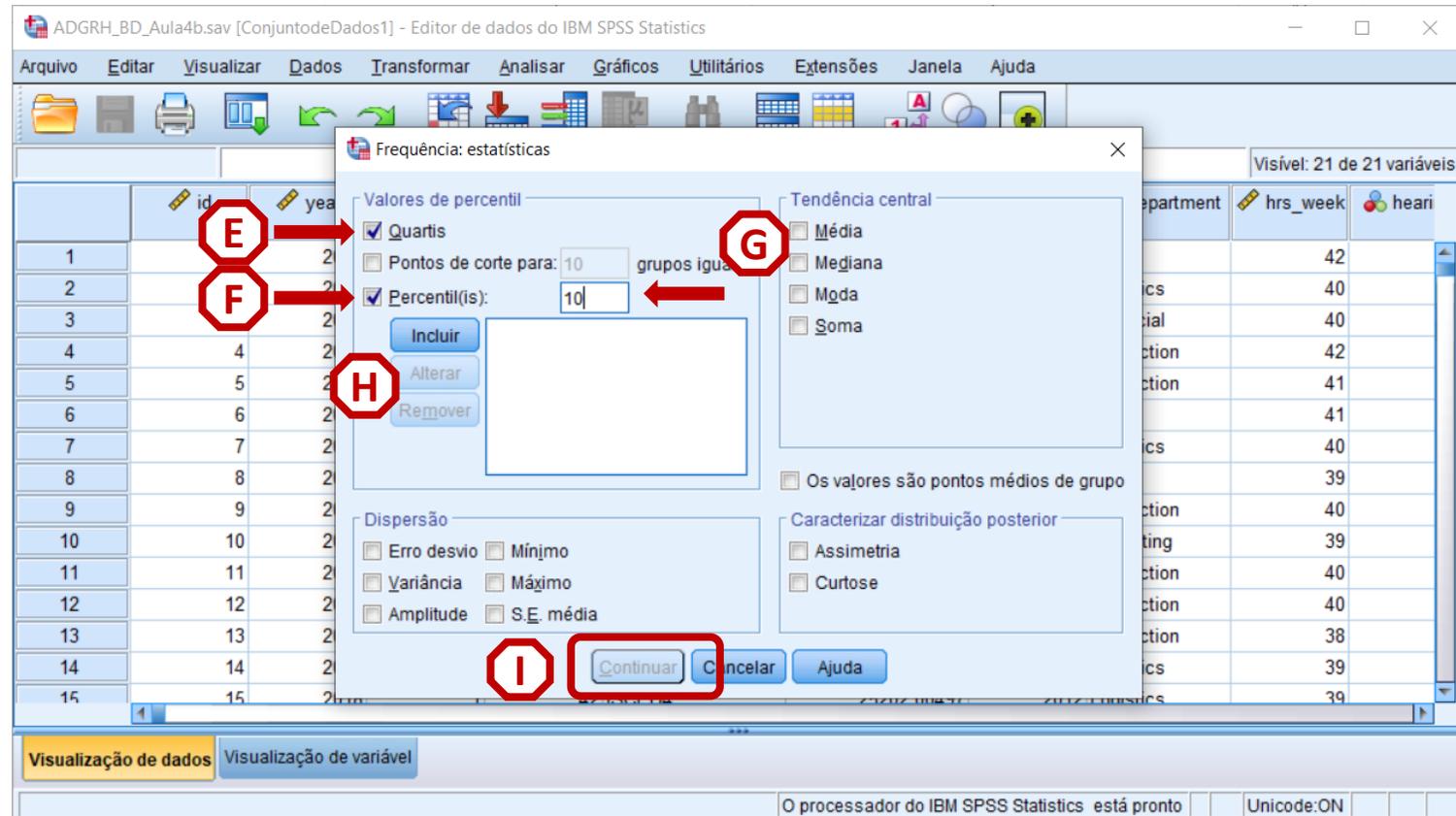


- Selecionar 'Incluir'



## Exercício: Inserir 9º Decil

- Selecionar '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 IBM SPSS Statistics 'Visualizador de Resultados' 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.

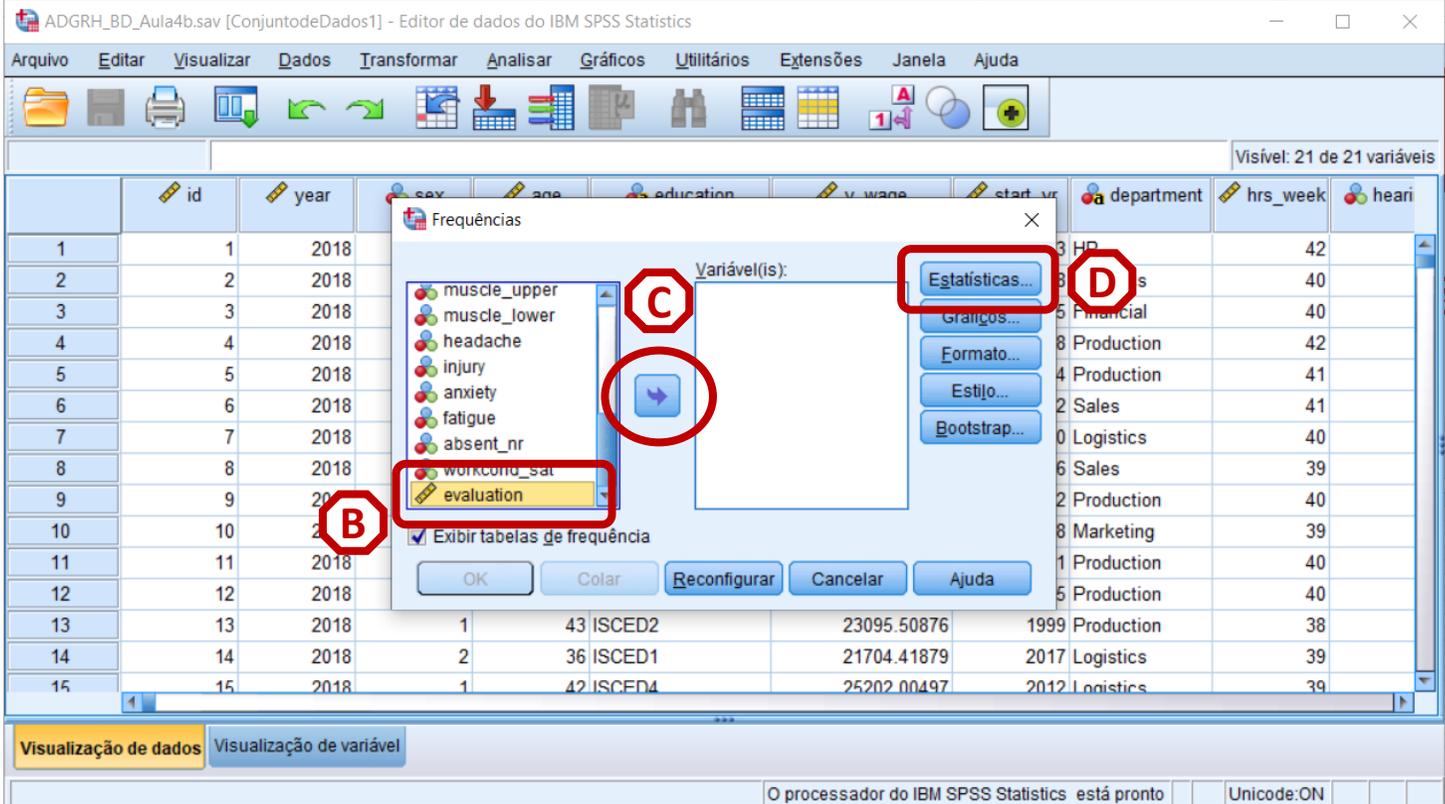
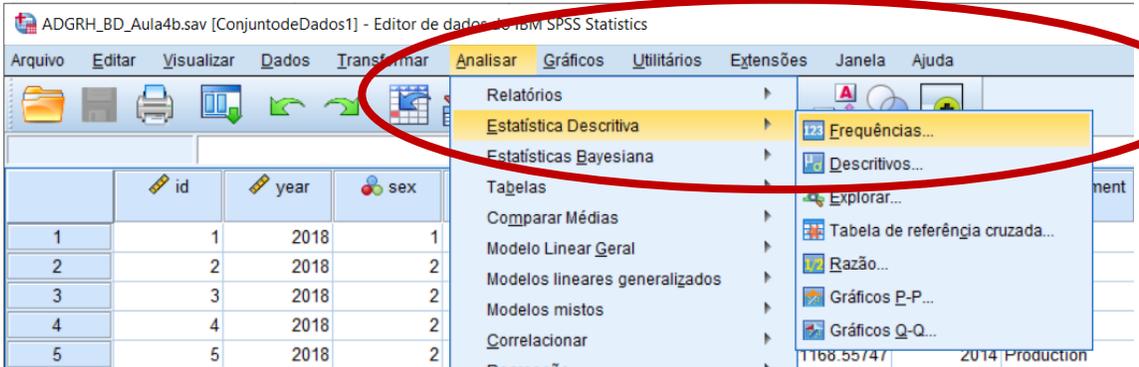
evaluation		
N	Válido	5000
	Omisso	0
Percentis	10	49.0805
	25	50.6639
	50	52.3968
	75	54.1703
	90	55.7755

	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	.3

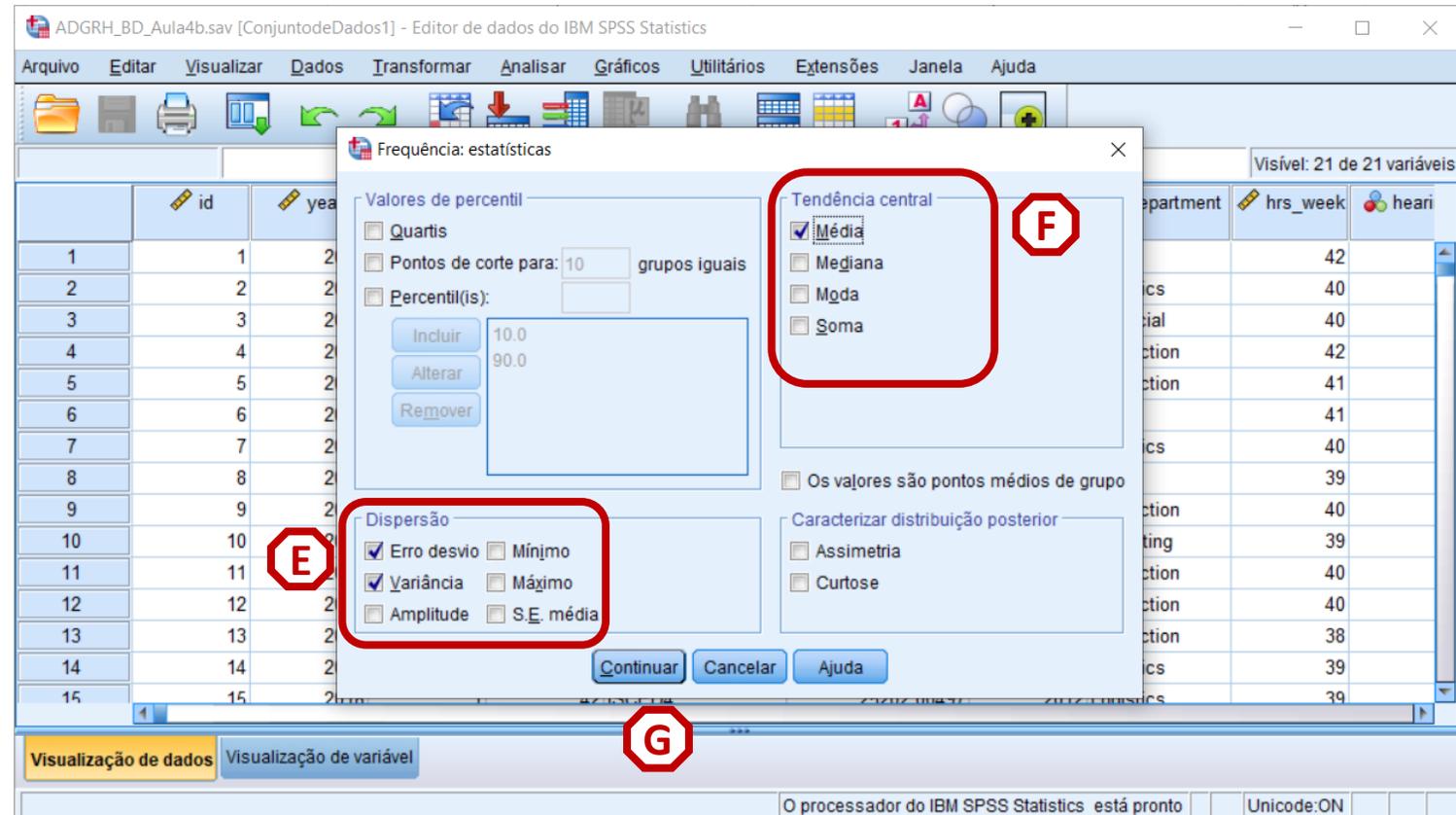
# Medidas de Dispersão

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



# Medidas de Dispersão

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



# Medidas de Dispersão

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

The screenshot shows the IBM SPSS Statistics Results Viewer window. The left pane displays a tree view of the results, with 'Frequências' selected for the variable 'evaluation'. The main pane shows the following output:

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

**Frequências**

**Estatísticas**

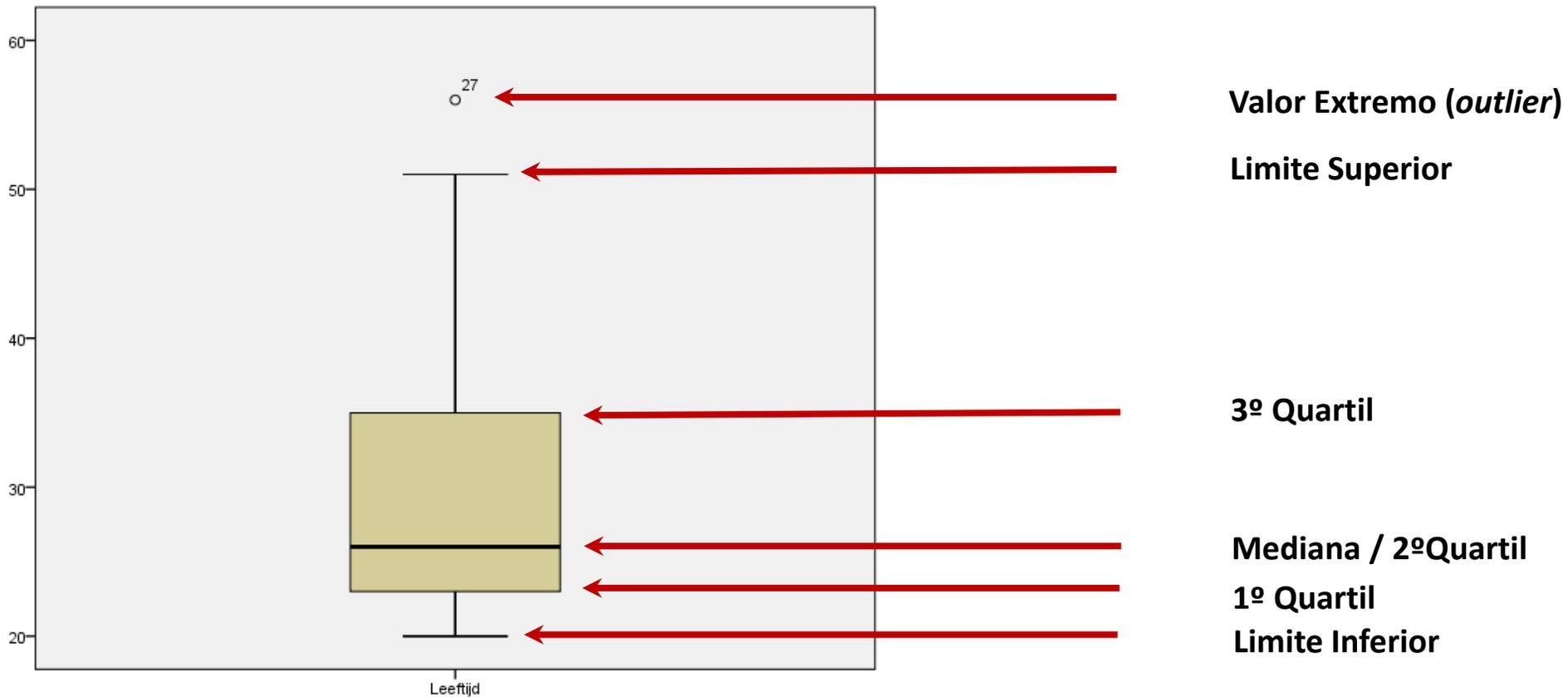
evaluation		
N	Válido	5000
	Omisso	0
Média		52.4194
Erro Desvio		2.61371
Variância		6.831

**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

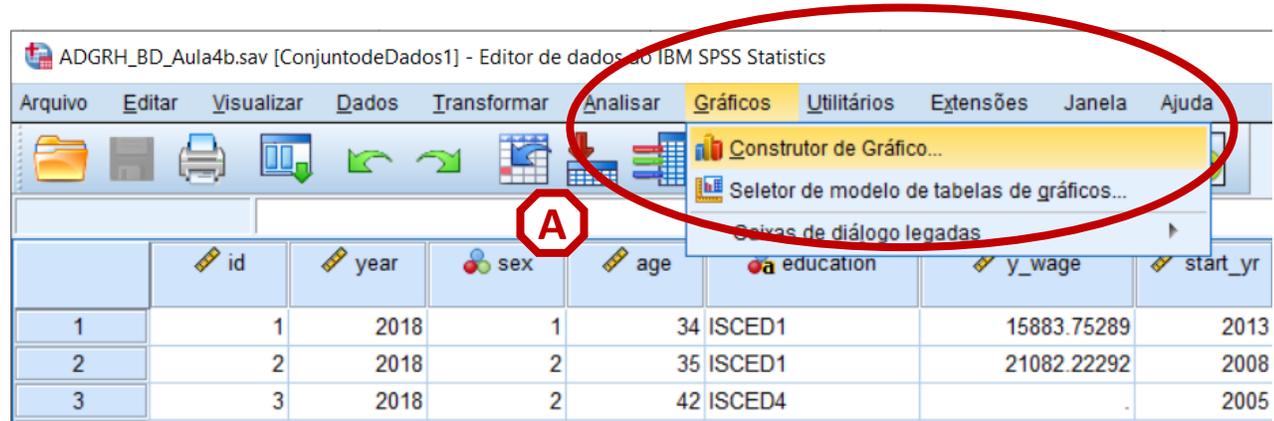
- Qual a melhor forma de visualizar o grau de dispersão da variável?

### Gráfico de Caixa



# Gráfico de Caixa

- Selecionar 'Gráficos' / 'Construtor de Gráfico'



# Gráfico de Caixa

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

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

Elementos básicos

Escolha entre:

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

Propriedades do elemento

Editar propriedades de:

Caixa1

X-Eixo1 (Caixa1)

Título 1

Estadísticas

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

OK Colar Reconfigurar Cancelar Ajuda

# 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’

**A** Variáveis:

**B** evaluation

**C** Galeria

**D** Escolha entre: Boxplot

**E** Eixo X?

Boxplot em 1-D

Filtro?

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de: Caixa1

X-Eixo1 (Caixa1)

Título 1

Estadísticas

Variável:

Estatística: Boxplot

Configurar parâmetros...

Exibir barra de erros

Representação de Barras de Erro

Intervalos de confiança

Nível (%): 95

Erro padrão

Multiplicador: 2

Desvio padrão

Multiplicador: 2

OK Colar Reconfigurar Cancelar Ajuda

# 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

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

Estadísticas

Variável:

Estatística:

Boxplot

Configurar parâmetros...

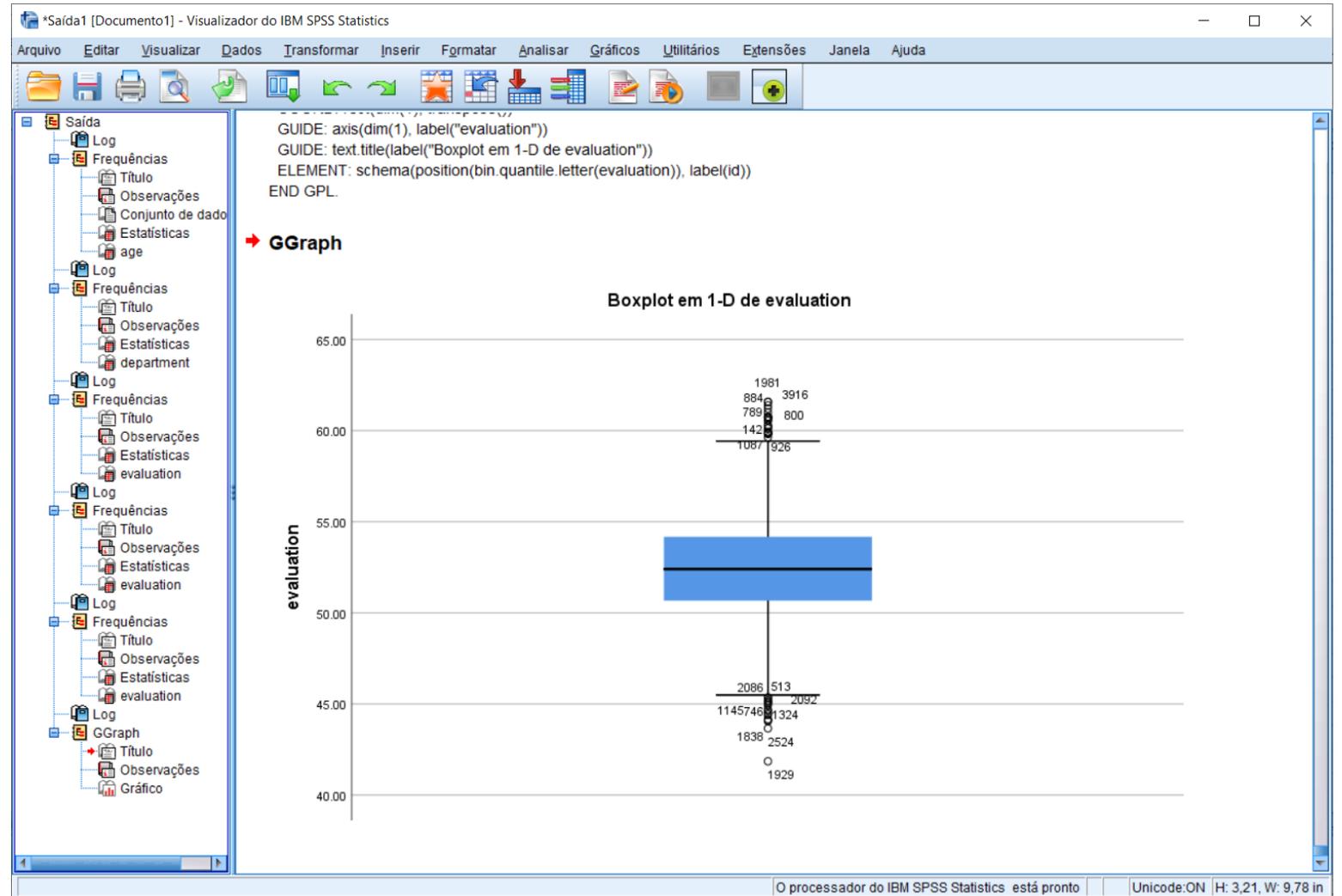
Exibir barra de erros

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'



## Limpar a Base de Dados

*Identificar/Recodificar 'Valores Omissos'*  
*Identificar/Recodificar 'Valores Extremos'*

## Limpar a Base de Dados

*Identificar/Recodificar 'Valores Omissos'*

# ANÁLISE DE DADOS EM GRH

## Aula 3: Identificar/Codificar Valores Omissos

- 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	.	9999999,00	.	.	.
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.**

- **Como é que podemos identificar ‘valores omissos numa base de dados?’**
  - **Em geral os valores omissos aparecem de três formas:**
    - **System Missing: ‘.’**
    - ***Sob a forma de texto: ‘Não Aplicável’***
    - ***Como um valor: ‘99’; ‘99999’***

- **Por que é que a existência de dados omissos é um problema?**
  - **Reduz o número de observações**
  - **Pode enviesar os resultados da nossa análise**

## Limpar a Base de Dados

*Identificar/Recodificar 'Valores Omissos'*

*Definir como Valor Omissos\**

*Recodificar como Valor Omissos*

*Substituir o Valor Omissos por Outro Valor (ex: Média)*

# Limpar a Base de Dados

*Identificar/Recodificar 'Valores Omissos'*  
*Definir 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.

GET  
FILE='C:\Users\amic\Google Drive\Professional\Teaching\MSc\MSc\_GRH\MSc\_GRH\_AD\2021\_22\Week\_3\ADGRH\_BD\_Aula3b.sav'.  
DATASET NAME ConjuntodeDados1 WINDOW=FRONT.  
FREQUENCIES VARIABLES=labforce  
/ORDER=ANALYSIS.

→ **Frequências**  
[ConjuntodeDados1] C:\Users\amic\Google Drive\Professional\Teaching\MSc\MSc\_GRH\MSc\_GRH\_AD\2021\_22\Week\_3\ADGRH\_BD\_Aula3b.sav

**Estadísticas** A  
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'



	Nome	Tipo	Largura	Decimais	Rótulo	Valores	Omissos	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, Employe...	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											

# 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, Employe...	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

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						Direito	Nominal	En
4	age	Numérico	8						Direito	Escala	En
5	marital	Numérico	8						Direito	Nominal	En
6	labforce	Numérico	8						Direito	Nominal	En
7	empstat	Numérico	8						Direito	Nominal	En
8	wage	Numérico	8						Direito	Escala	En
9	hearing	Numérico	8						Direito	Nominal	En
10	eye	Numérico	8						Direito	Nominal	En
11	memory	Numérico	8						Direito	Nominal	En
12	physical	Numérico	8						Direito	Nominal	En
13	mobility	Numérico	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'
- Selecionar 'Valores omissos discretos', e instruir o SPSS que o valor 1 deve ser tratado como valor omissos.
- Clicar 'OK'



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						Direito	Nominal	En
4	age	Numérico	8						Direito	Escala	En
5	marital	Numérico	8						Direito	Nominal	En
6	labforce	Numérico	8						Direito	Nominal	En
7	empstat	Numérico	8						Direito	Nominal	En
8	wage	Numérico	8						Direito	Escala	En
9	hearing	Numérico	8						Direito	Nominal	En
10	eye	Numérico	8						Direito	Nominal	En
11	memory	Numérico	8						Direito	Nominal	En
12	physical	Numérico	8						Direito	Nominal	En
13	mobility	Numérico	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'
- Selecionar 'Valores omissos discretos', e instruir o SPSS que o valor 1 deve ser tratado como valor omissos.
- Clicar 'OK'
- Clicar o ícone 'Salvar este documento'



\*ADGRH\_BD\_Aula3b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

		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... 1		8	Direito	Nominal	En
7	empstat	Numérico	8	0	Employment St...	{1, Employe...	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



Eu não aconselho este método!!!

Recodificar para uma variável diferente é uma opção muito mais segura, pois permite-nos manter a variável original!

## Limpar a Base de Dados

*Identificar/Recodificar 'Valores Omissos'*  
*Recodificar como Valor Omisso*

# 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	107745	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'



\*ADGRH\_BD\_Aula3b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janelas

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

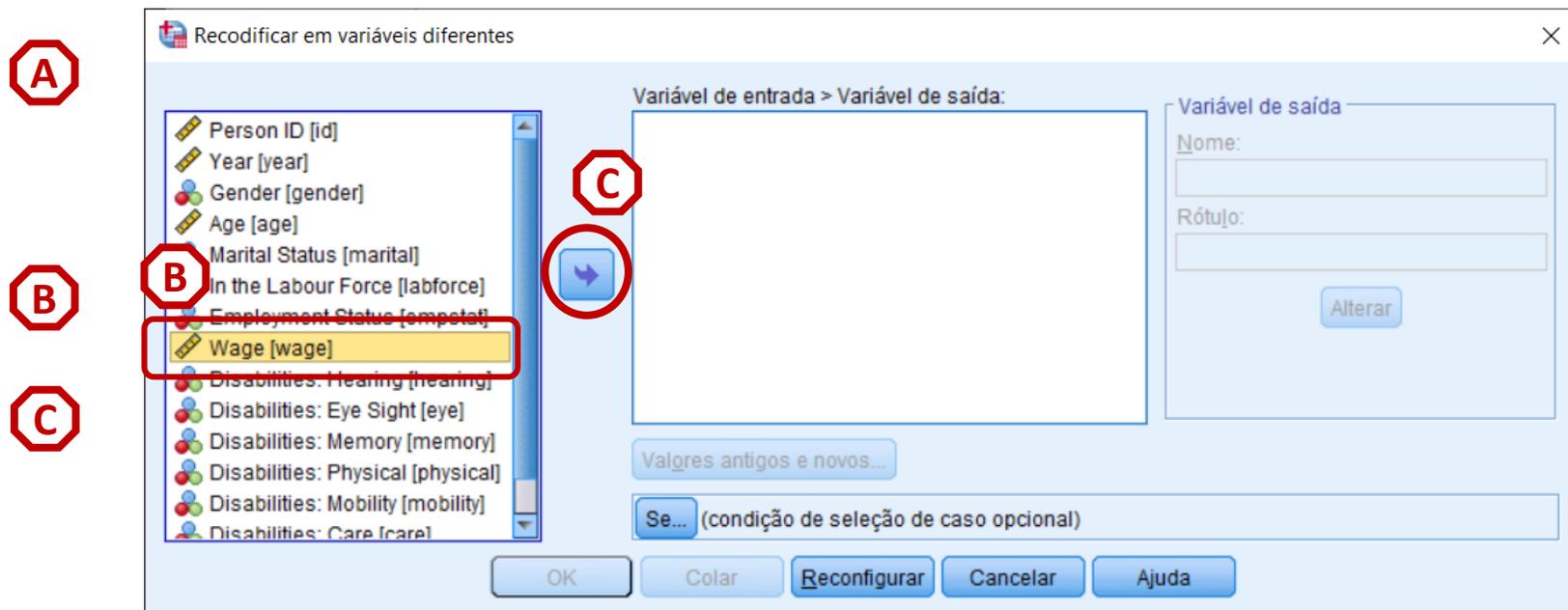
	id	year
1	126962	2015
2	16717	2015
3	22360	2015
4	23780	2015
5	134027	2015

empstat

4
1
1
4
.

# 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

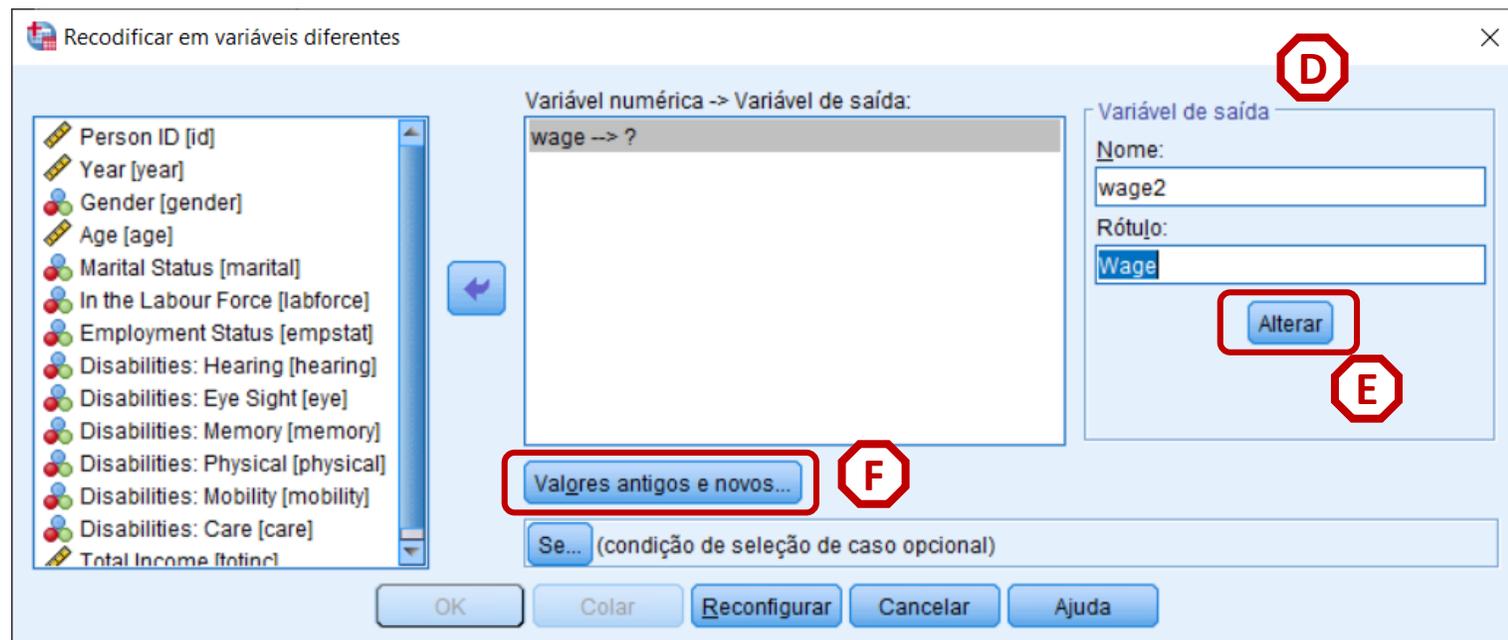
B

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 seqüências de caracteres Largura: 8

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

Incluir

Alterar

Remover

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'



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 seqüências de caracteres Largura:

Converter seqüências 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'

G

H

I

J

K

L

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: 8

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

Continuar Cancelar Ajuda

# Recodificar como Valor Omissso

- 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:

Omissso no sistema

Omissso 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:

Omissso 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

- Let's check if the variable was adequately recoded

\*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

*Identificar/Recodificar 'Valores Omissos'*

*Substituir o Valor Omissos por Outro Valor (ex: Média)*

# Substituir o Valor Omissos 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	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

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

mobility care

1	1
1	1
1	1
.	.
.	.
1	1
1	1
.	.
1	1
1	1
1	1
1	1
.	.
1	1

Substituir valores omissos...

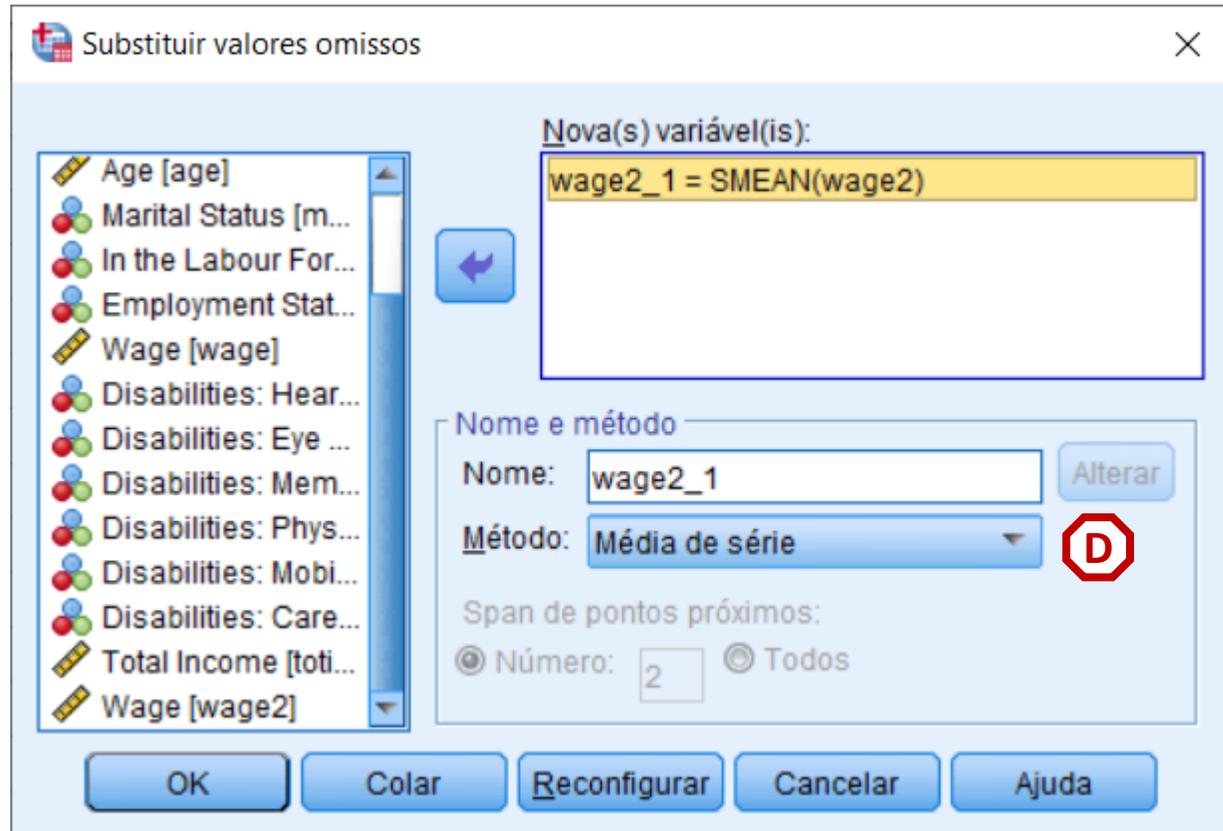
# Substituir o Valor Omisso pela Média

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



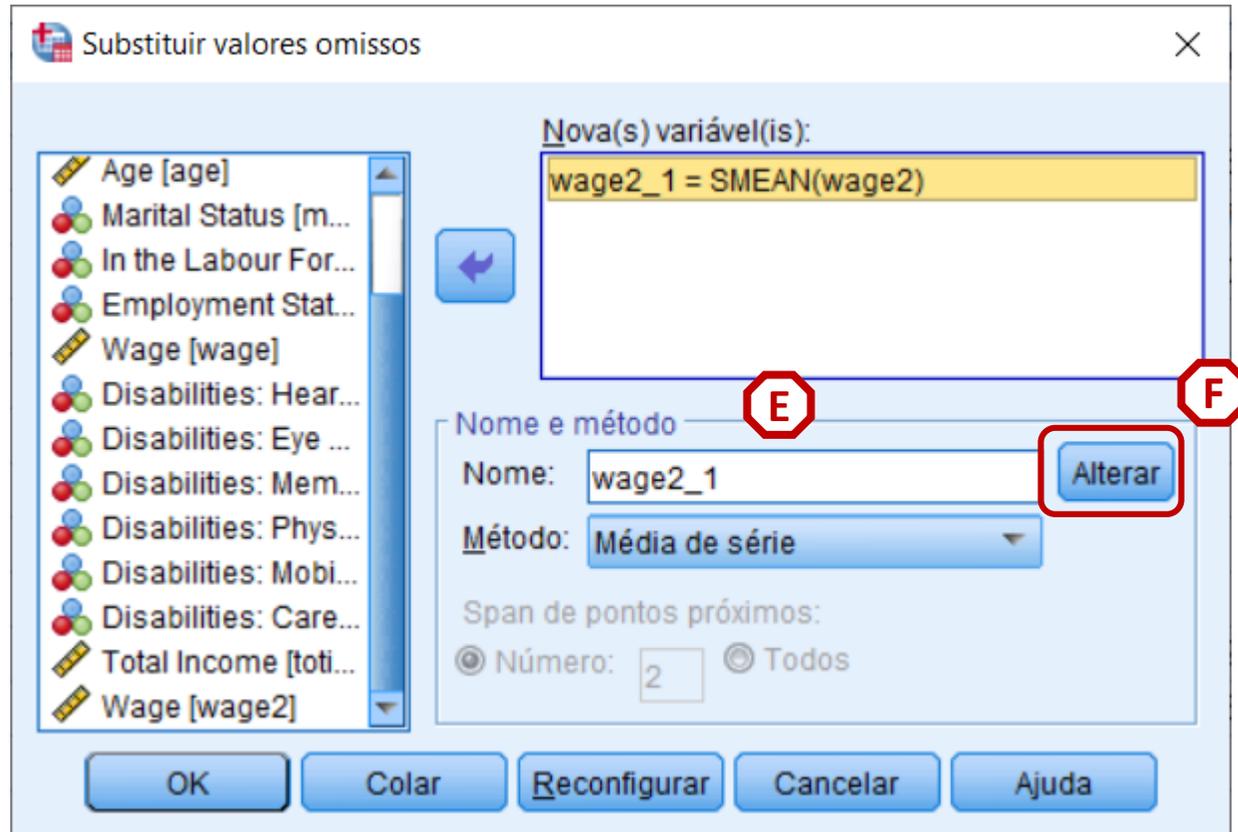
# Substituir o Valor Omissos pela Média

- 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**
- Change the name of the new variable to 'wage3' **E**
- Clicar 'Alterar' **F**



# Substituir o Valor Omisso pela Média

- Selecionar 'Transformar' / 'Replace Missing Values'
- Selecionar the 'wage2' variable
- Place it in the 'New Variable(s)' box
- By default the method of replacement is 'Series mean'
- Change the name of the new variable to 'wage3'
- Clicar 'Alterar'
- Clicar 'OK'

A

B

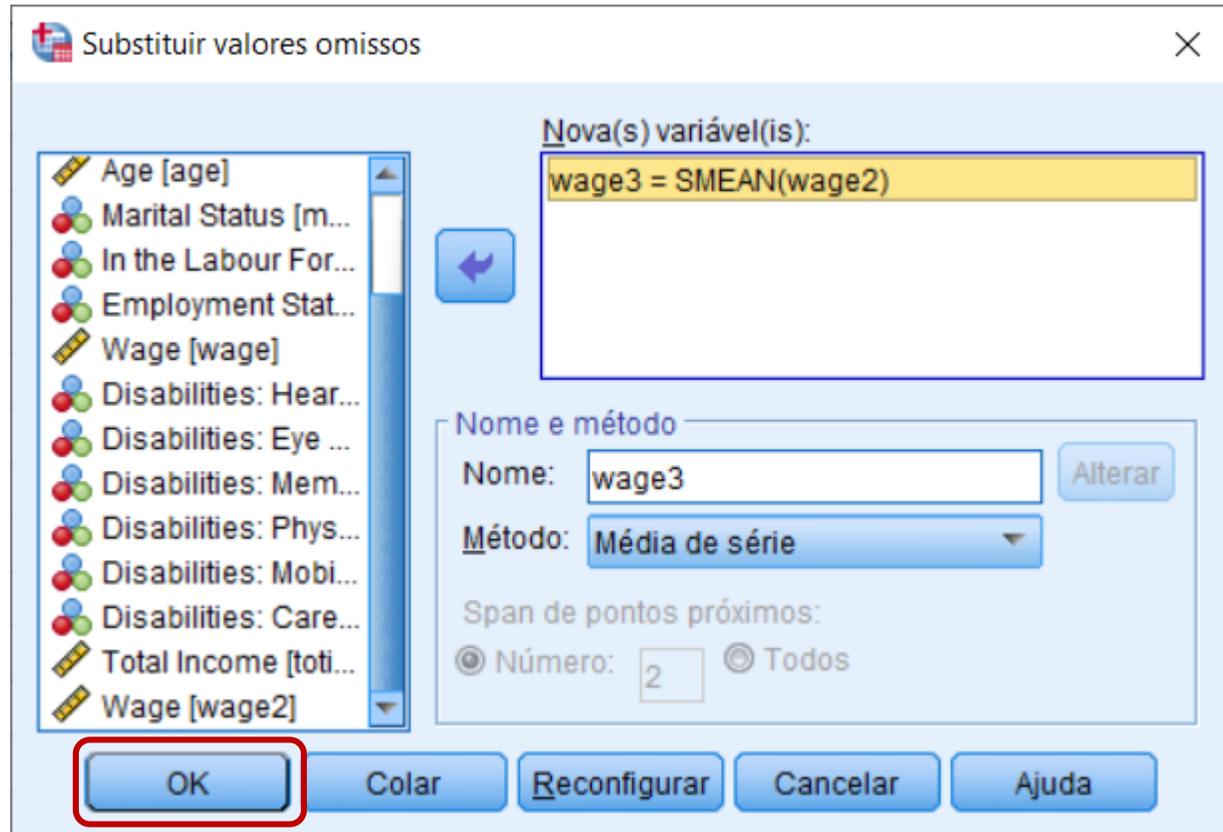
C

D

E

F

G



G

# Substituir o Valor Omisso pela Média

- Como se pode ver, 2806 valores omisso foram recodificados...

... o que significa que na variável wage 3 todas as 5000 observações têm dados sobre o salário

A

B

**Estadística Descritiva**

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

→ **Substituir valores omisso**

```
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 Omisso Substituídos	Número de Caso de Valores Não omisso		N de Casos Válidos	Criando Função
			Primeiro	Último		
1	wage3	2806	1	5000	5000	SMEAN (wage2)

**Descritivos**

**Estadística Descritiva**

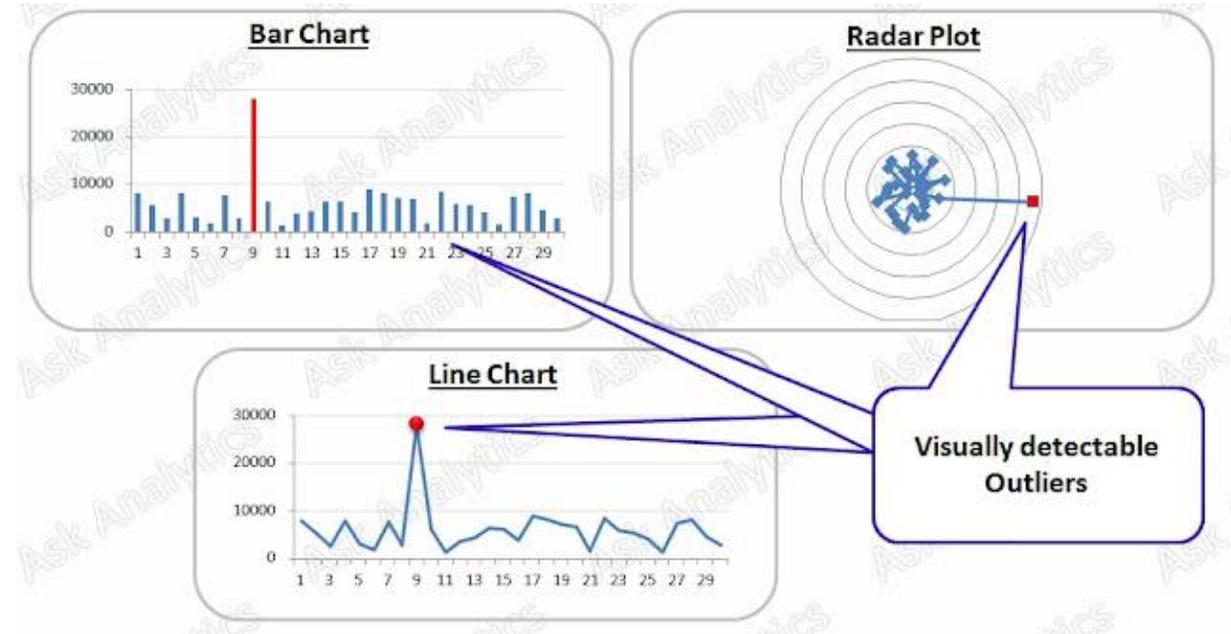
	N	Mínimo	Máximo	Média	Erro Desvio
SMEAN(wage2)	5000	10,00	250000,00	42572,8888	24495,95134
N válido (de lista)	5000				

B

## Limpar a Base de Dados

*Identificar/Recodificar 'Valores Extremos'*

- O que são ‘valores extremos’?
  - São observações que se desviam de uma forma muito significativa do padrão de distribuição das observações de uma variável.
  - Só acontecem em variáveis razão.



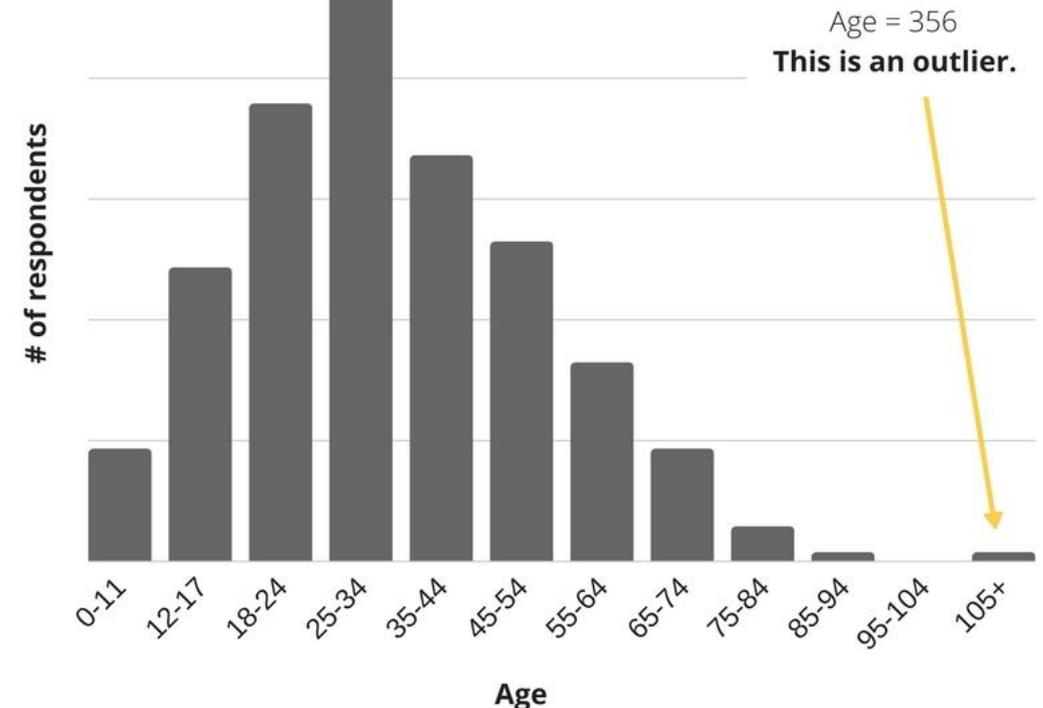
Fonte: <http://www.askanalytics.in/p/outl.html>

- **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) \***

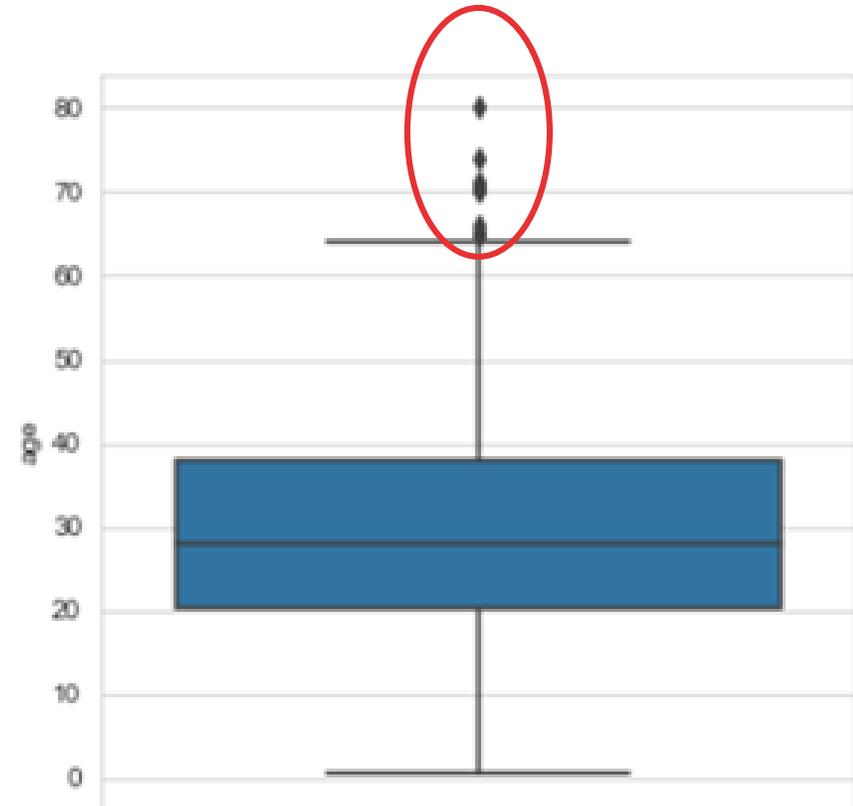
\* Nos casos em que valores omissos codificados como '99999' (ou outras sequências), estes não devem ser tratados como valores extremos.

- **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 da Possibilidade de 'Valores Extremos'*

*Usar Z-Scores para Determinar 'Valores Extremos'*

*Inspeccionar Potenciais 'Valores Extremos'*

*Recodificar como Valor Omisso*

## Limpar a Base de Dados

*Identificar/Recodificar 'Valores Extremos'*

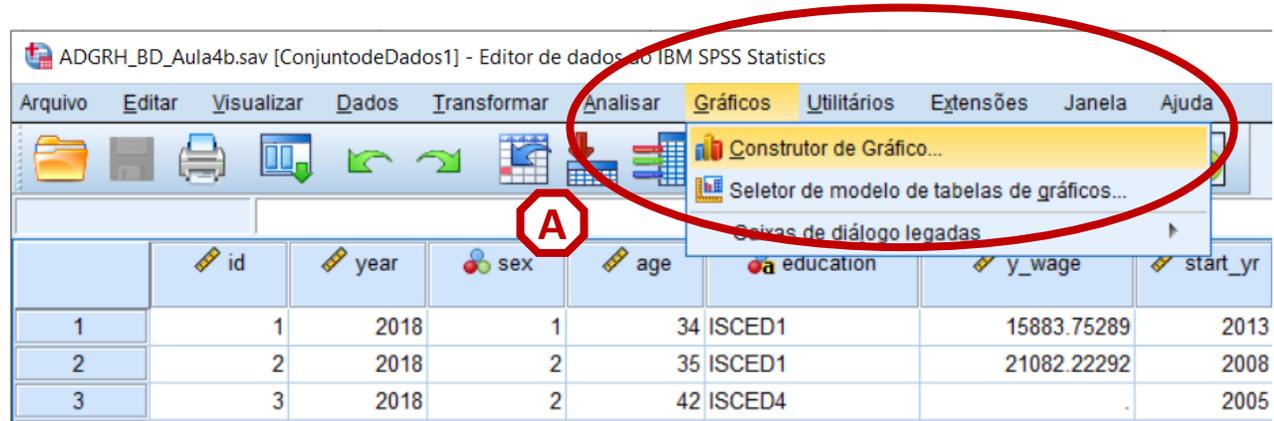
*Diagnóstico Visual da Possibilidade de 'Valores Extremos'*

- **Objectivo:**
  - **Perceber se existem valores extremos na variável que mede o salário anual dos trabalhadores?**

# 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)

The screenshot shows the 'Construtor de gráfico' window. On the left, a list of variables includes 'id', 'year', 'sex', 'age', 'education', 'y\_wage', 'start\_yr', 'department', 'hrs\_week', and 'hearino'. The central plot area displays a 'Boxplot em 1-D' with a blue box and whiskers. A dashed box labeled 'Eixo X?' is on the left, and another labeled 'Filtro?' is on the right. Below the plot, the 'Galeria' tab is active, showing a list of chart types: 'Favoritos', 'Barra', 'Linha', 'Área', 'Setor/Polar', 'Dispersão/Ponto', 'Histograma', 'Mapa de Pontos', 'Boxplot', and 'Eixos duplos'. The 'Boxplot' option is highlighted with a red box labeled 'B'. In the 'Escolha entre:' section, a simple boxplot icon is highlighted with a red box labeled 'C'. The right panel shows the 'Propriedades do elemento' tab, with 'Caixa1' selected. The 'Estatísticas' section shows 'Variável:' and 'Estatística:' set to 'Boxplot'. The 'Representação de Barras de Erros' section has 'Intervalos de confiança' selected with a 95% level and a multiplier of 2. At the bottom, there are buttons for 'OK', 'Colar', 'Reconfigurar', 'Cancelar', and 'Ajuda'.

# Gráfico de Caixa

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

A

B

C

D

E

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- year
- sex
- age
- education
- start\_yr
- start\_age
- experience
- y\_wage
- department
- week

uma categoria (variável 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

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

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

OK Colar Reconfigurar Cancelar Ajuda

# Gráfico de Caixa

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

A

B

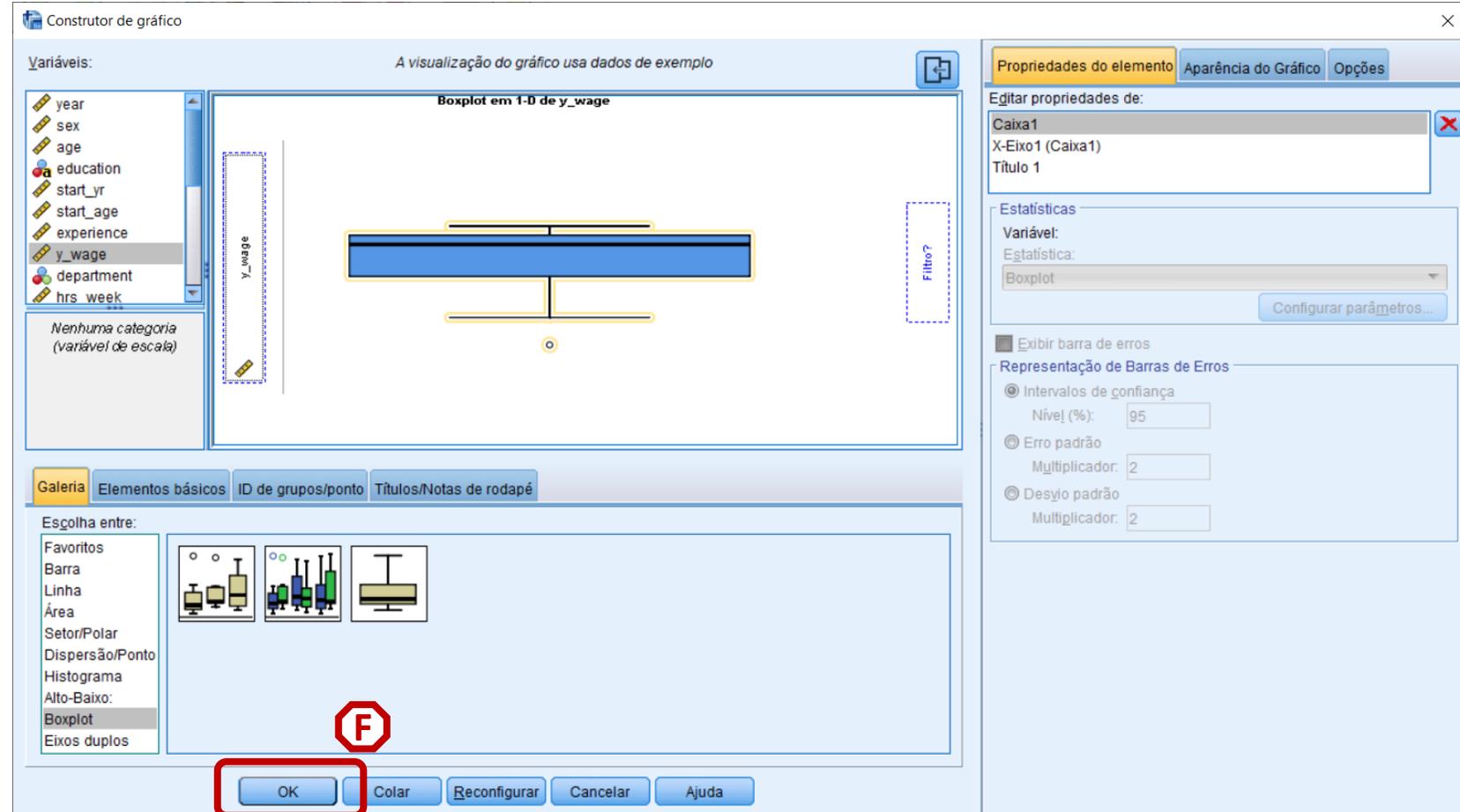
C

D

E

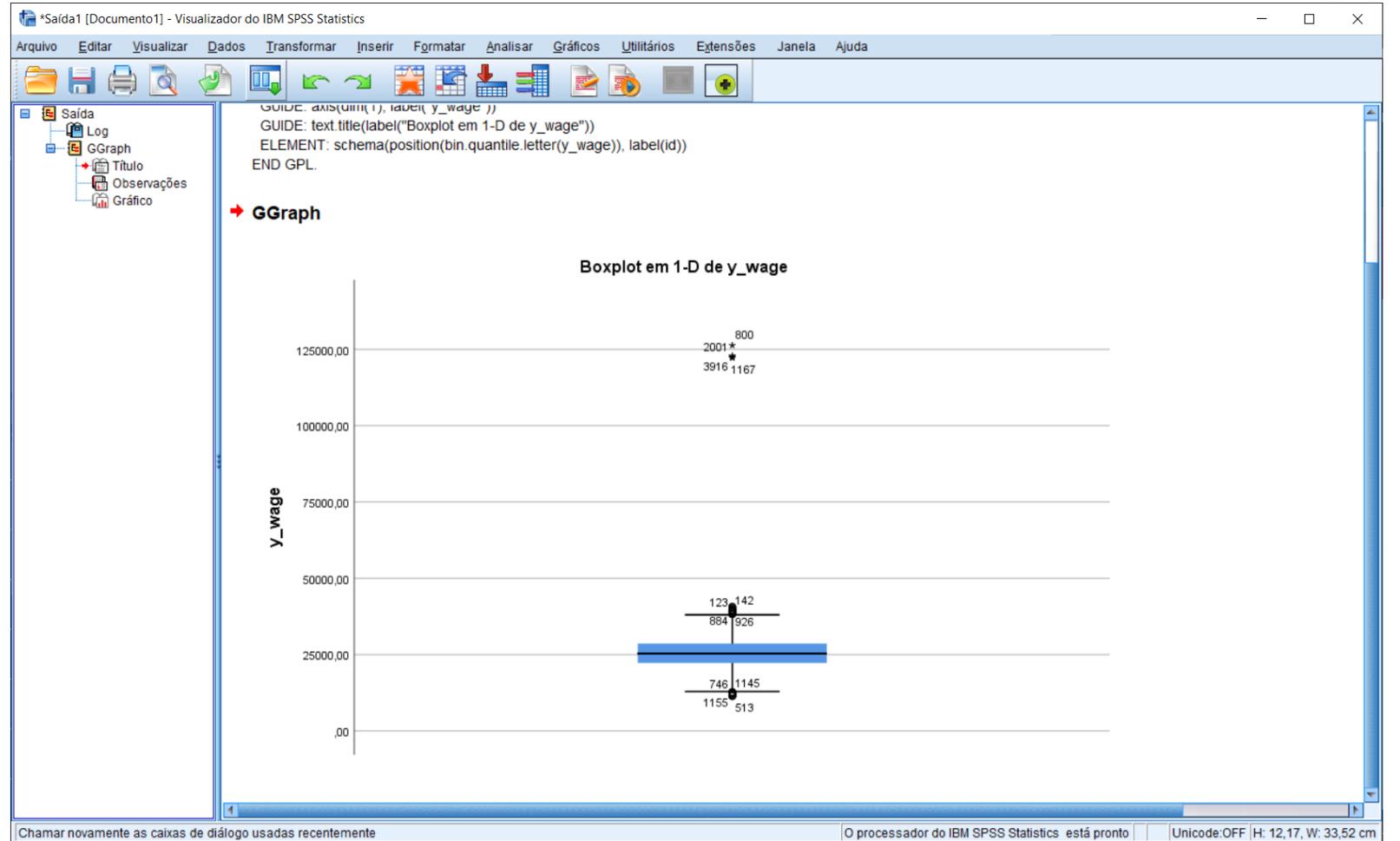
F

F



# Gráfico de Caixa

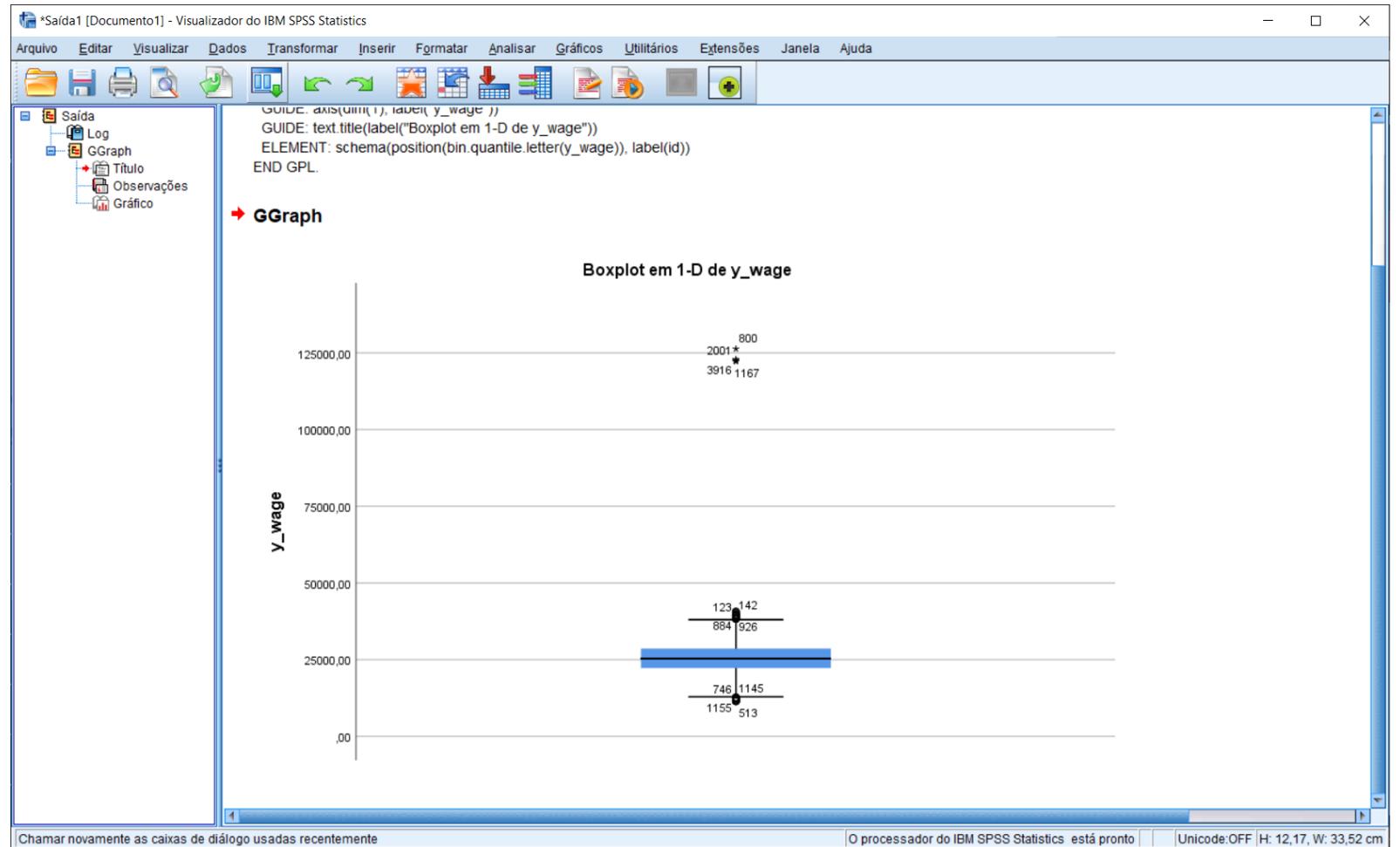
- O gráfico é publicado no 'Visualizador de Resultados'



# Gráfico de Caixa

- O gráfico é publicado no 'Visualizador de Resultados'

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



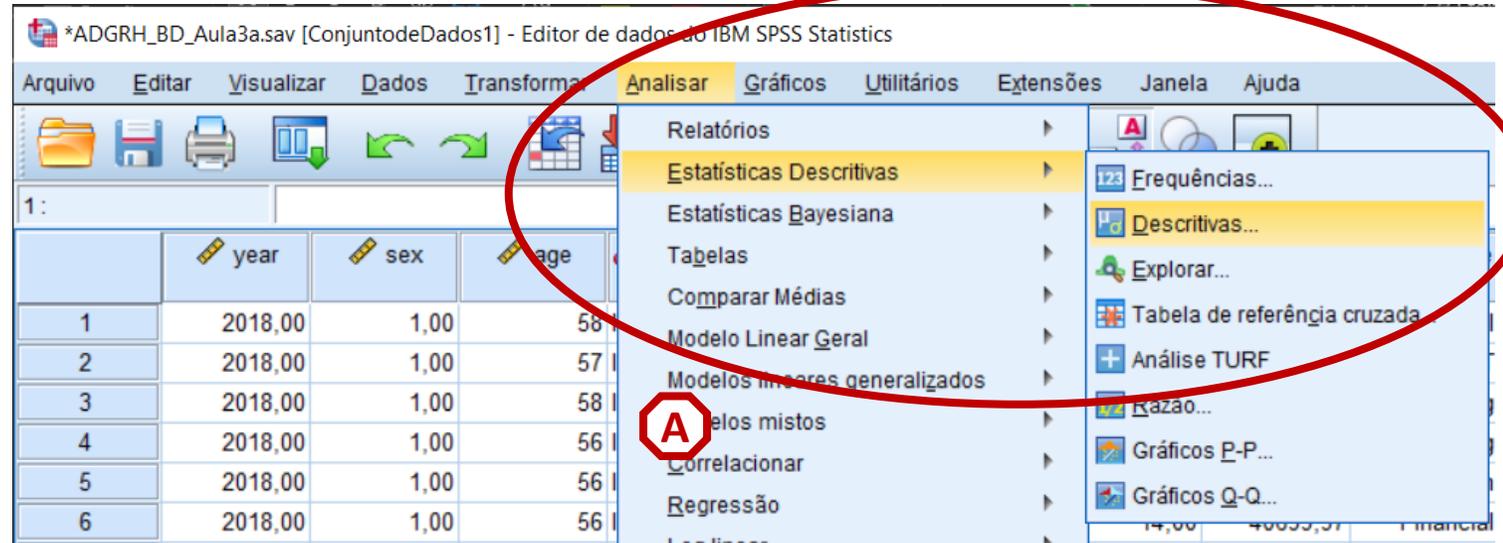
## Limpar a Base de Dados

*Identificar/Recodificar 'Valores Extremos'*

*Usar Z-Scores para Determinar 'Valores Extremos'*

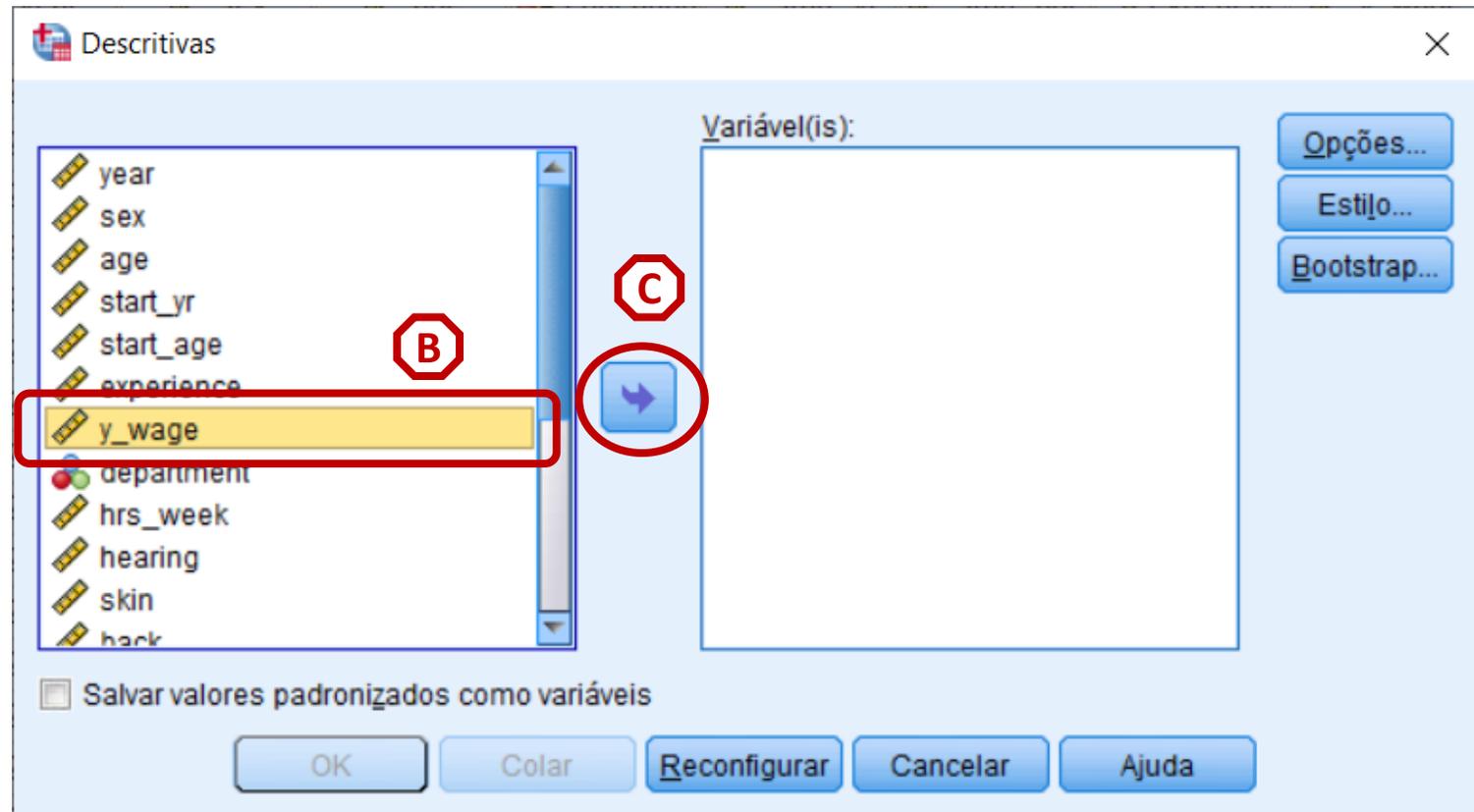
# Z-Scores para identificar outliers

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



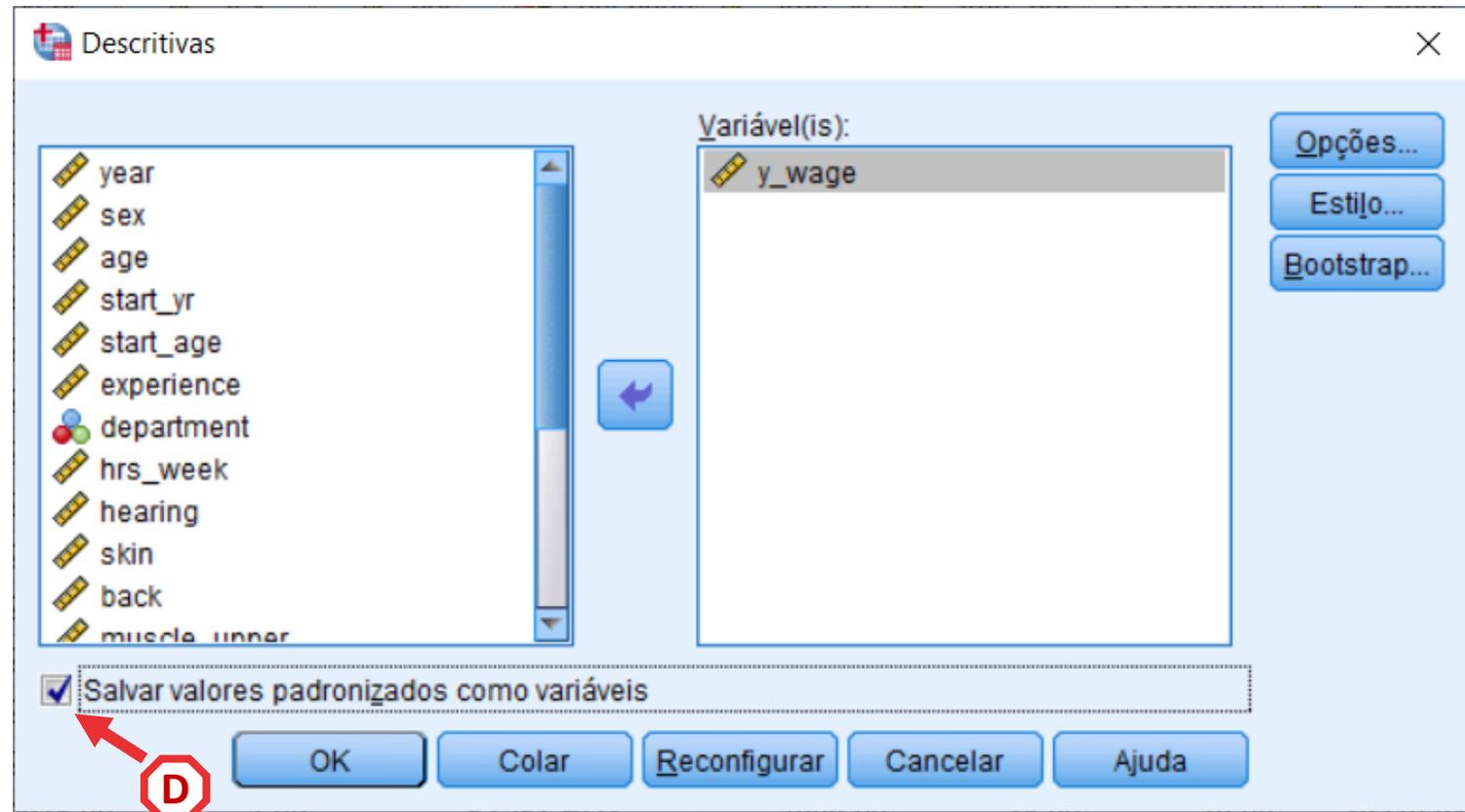
# Z-Scores para identificar outliers

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Descritivas' A
- Selecionar a variável 'y\_wage' B
- Colocar na caixa 'Variável(is)' C



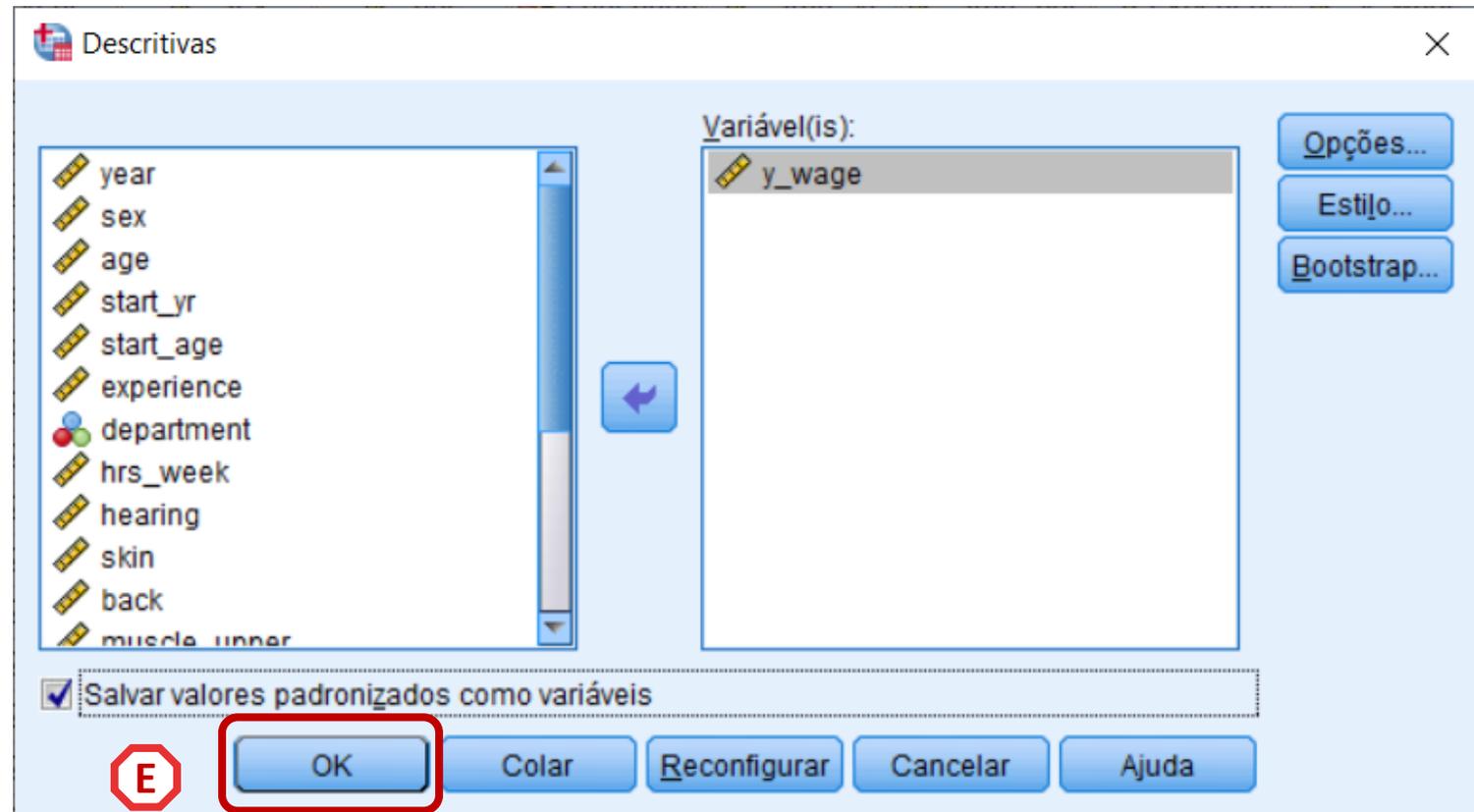
# Z-Scores para identificar outliers

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Descritivas' ⓐ
- Selecionar a variável 'y\_wage' ⓑ
- Colocar na caixa 'Variável(is)' ⓒ
- Seleccionar a opção 'Salvar valores padronizados como variáveis' (z-scores) ⓓ



# Z-Scores para identificar outliers

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Descritivas' ⓐ
- Selecionar a variável 'y\_wage' ⓑ
- Colocar na caixa 'Variável(is)' ⓒ
- Seleccionar a opção 'Salvar valores padronizados como variáveis' (z-scores) ⓓ
- Clicar 'OK' / 'Continuar' ⓔ



# 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	54,14	1,00263				
7	,00	,00	,00	6	Not Very S...	49,65	-.82585				
8	,00	1,00	,00	10	Not At All ...	46,64	-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	54,64	,49816				
13	,00	,00	,00	13	Not Very S...	56,64	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'



\*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	-				
2	,00	,00	,00	10	Satisfied	52,58	-				
3	,00	,00	,00	15	Satisfied	53,80	-				
4	,00	,00	,00	10	Satisfied	50,42	-				
5	,00	,00	,00	11	Not At All ...	48,43	-1				
6	,00	,00	,00	10	Satisfied	54,34	1				
7	,00	,00	,00	6	Not Very S...	49,65	-				
8	,00	1,00	,00	10	Not At All ...	46,84	-1				
9	,00	,00	,00	8	Satisfied	50,76	-				
10	,00	,00	,00	17	Not Very S...	53,33	-				
11	,00	,00	,00	14	Satisfied	50,59	-				
12	,00	,00	,00	7	Satisfied	54,64	-				
13	,00	,00	,00	13	Not Very S...	56,64	1				
14	,00	,00	,00	7	Not Very S...	55,36	0,0239				
15	,00	,00	,00	11	Not Very S...	56,09	1,17733				

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:
- 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...	61,66	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

- $< -3.28$



\*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...	61,66	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

*Identificar/Recodificar 'Valores Extremos'*  
*Inspeccionar Potenciais 'Valores Extremos'*

# 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

146

## Limpar a Base de Dados

*Identificar/Recodificar 'Valores Extremos'*  
*Recodificar como Valor Omisso*

## 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	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	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

148

## Recodificar como Valor Omissos

- 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...	60,72	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	

# Recodificar como Valor Omisso

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



\*ADGRH\_BD\_Aula3b.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janelas

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

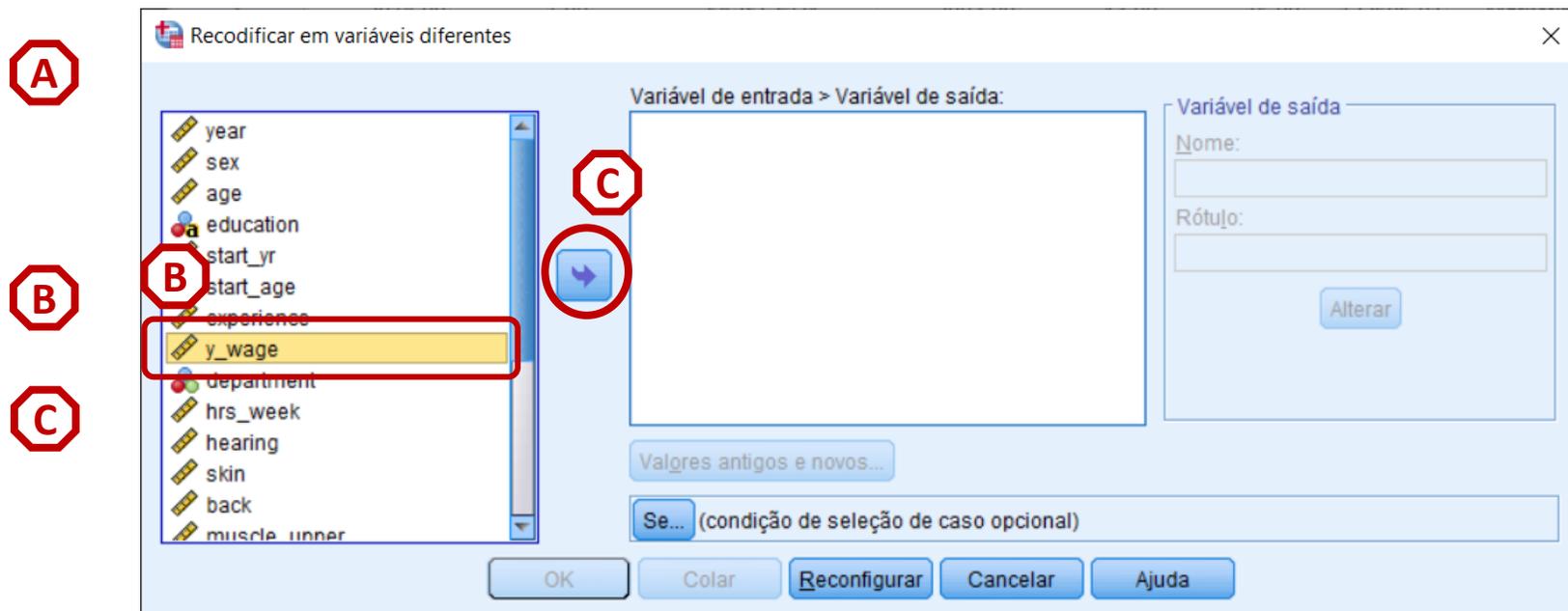
	id	year
1	126962	2015
2	16717	2015
3	22360	2015
4	23780	2015
5	134027	2015

empstat

4
1
1
4
.

# 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\_wage2' e o rótulo como 'Wage'
- Clicar 'Alterar'
- Clicar e 'Valores antigos e novos'

A

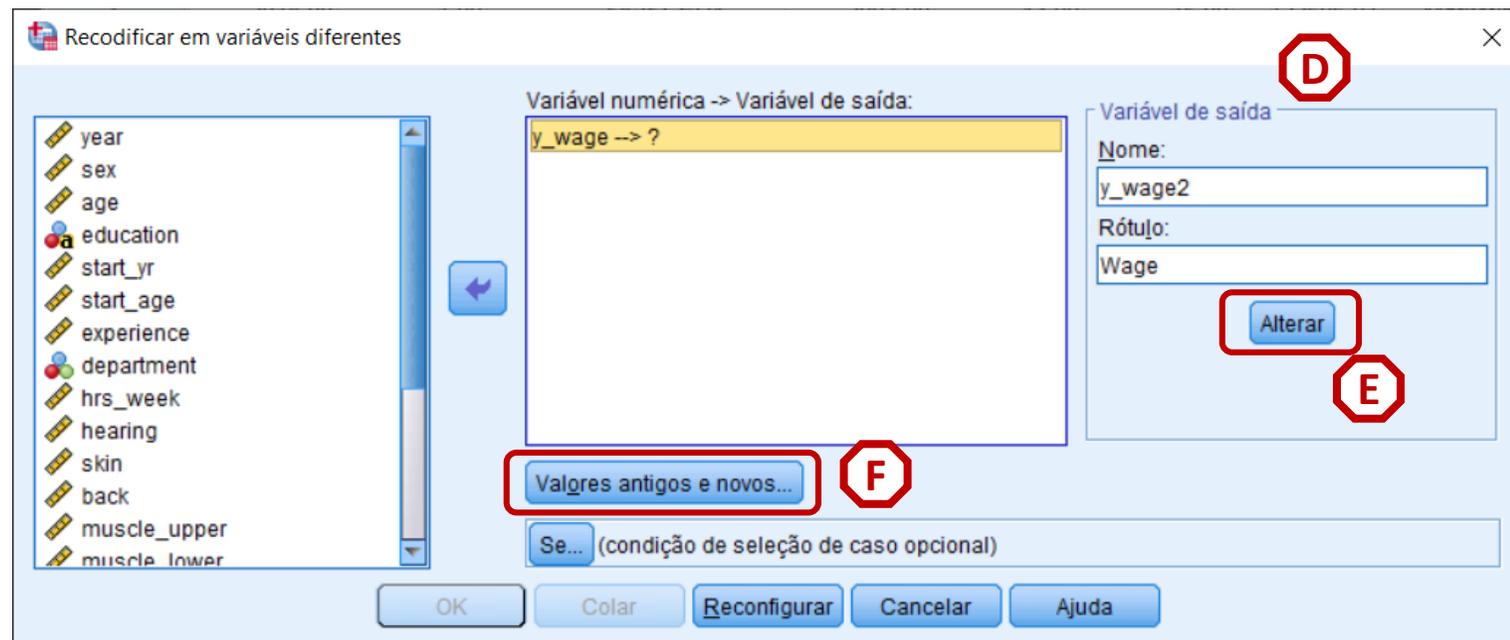
B

C

D

E

F



# Recodificar como Valor Omissso

- 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'



Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo

Valor:

Omissso no sistema

Omissso para o sistema ou usuário

Amplitude:

até

Intervalo, MAIS BAIXO até o valor:

Intervalo, valor até o MAIS ALTO:

40700

Todos os outros valores

Novo valor

Valor:

Omissso 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

# 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 seqüências de caracteres Largura: 8

Converter seqüências 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'

G

H

I

J

K

L

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 Omissso

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

G

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

H

- Clicar 'Incluir'

I

- Depois, definir que 'Todos os outros valores'

J

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

K

- Clicar 'Incluir'

L

- Clicar 'Continuar' / 'OK'

M

Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo

Valor:

Omissso no sistema

Omissso 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:

Omissso 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: 8

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

Continuar Cancelar Ajuda

# Recodificar como Valor Omissão

- Vamos ver se a operação de recodificação foi bem sucedida

\*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_wage2	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

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 [Facebook] [Twitter] [Google+] [LinkedIn] [Email]



Int qu

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.

horários visitas

A NOVA NET DA NOS É GIGA

Saber mais

## ÚLTIMAS

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

# Até Para a Semana!