



## *Aula 9: Que factores explicam a variação nos salários na organização?*

*O modelo de regressão linear*

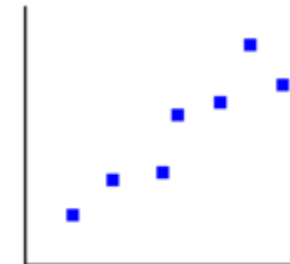
**Docente:** Amílcar Moreira

**Data & Hora:** 24/11/2020, 20:30-22:30h

**Local:** FRANCESINHAS 2, Sala 101

- **Na Aulas Anteriores**
  - Exploramos as bases da estatística inferencial
- **Objetivos da Aula**
  - **Parte Teórica**
    - O modelo de regressão linear como base da análise estatística multivariada, de carácter inferencial
    - Saber descrever um modelo analítico através da equação do modelo de regressão linear
  - **Parte Prática**
    - Saber implementar o modelo de regressão linear
    - Saber interpretar os resultados do modelo de regressão linear

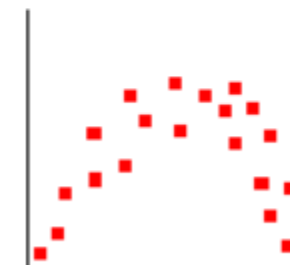
- **Coeficientes de Associação / Correlação**
  - Permitem aferir se existe uma relação sistemática entre duas variáveis
  - Permitem aferir a força da relação sistemática entre duas variáveis
  - Permitem identificar a direção da relação sistemática entre duas variáveis (correlação apenas)
- **Modelo de regressão linear (simples)**
  - Diferencia entre:
    - Variável Dependente (DV)
    - Variável Independente (IV), que influencia a variável dependente
    - Presume a existência uma relação linear entre as duas variáveis
  - O modelo de regressão pode depois ser alargado para incluir de mais do que uma variável independente



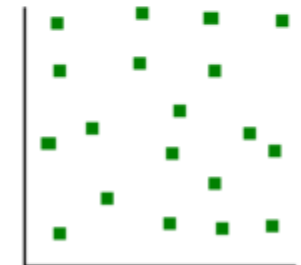
Positive linear relationship



Negative linear relationship



Non-linear relationship



No relationship

- **Modelo de regressão linear (simples)**

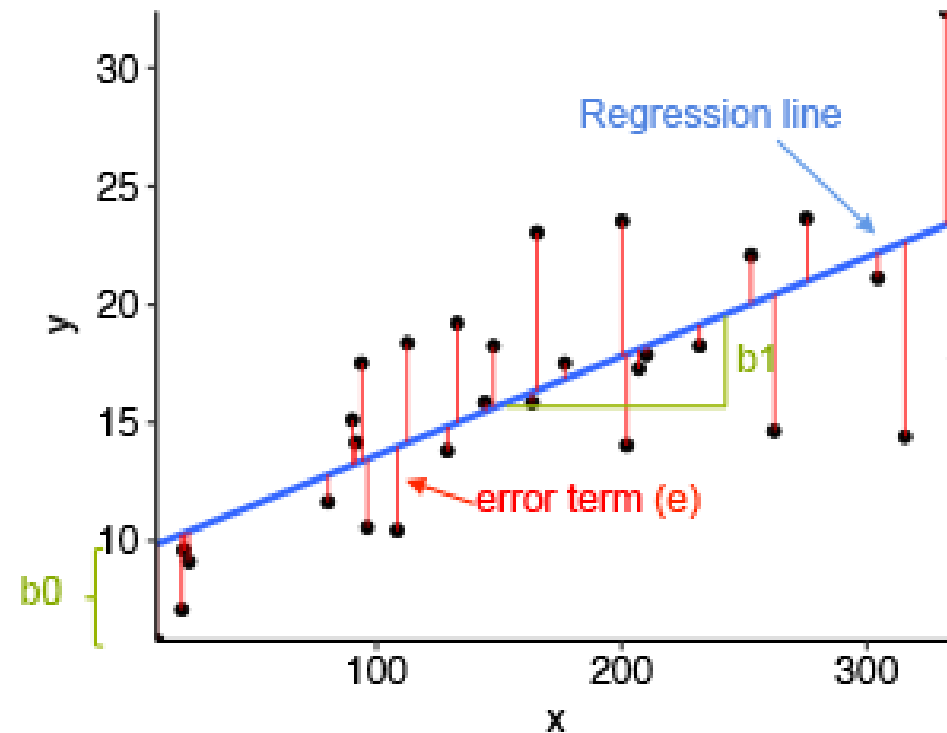
$$Y = \beta_0 + \beta_1 X + E$$

$Y$  Variável Dependente

$\beta_0$  Constante (i.e., valor de  $Y$  quando  $X = 0$ )

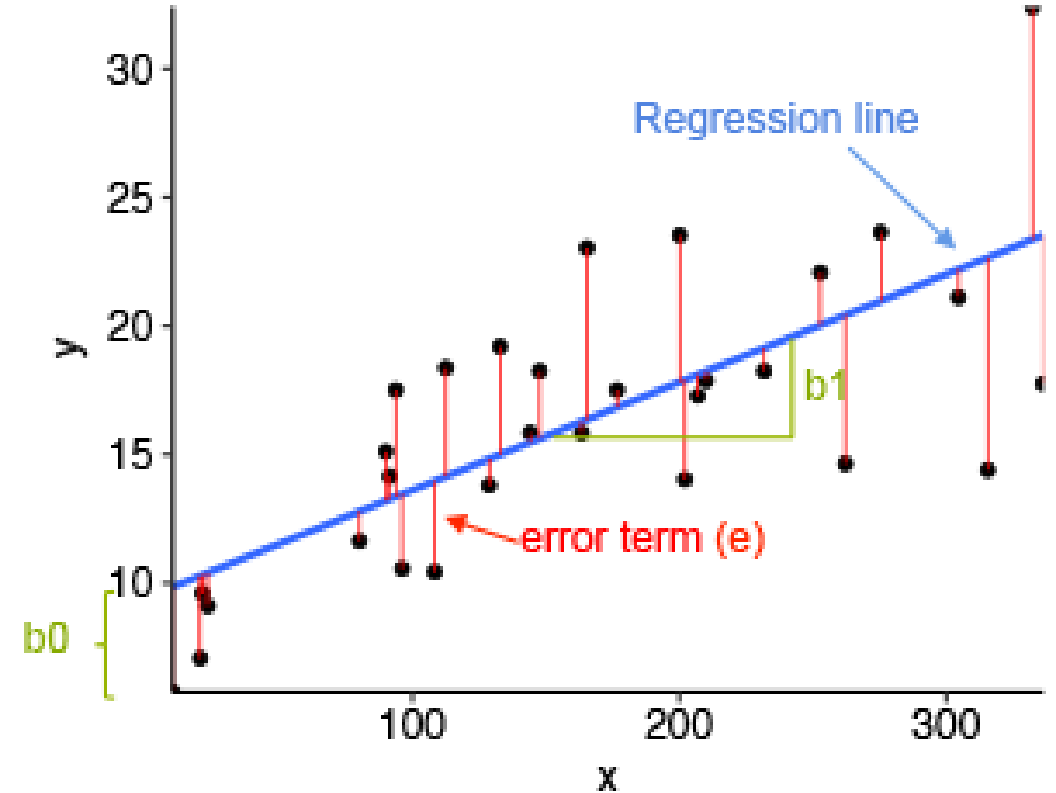
$\beta_1 X$  Coeficiente Beta da variável  $X$  (mede o efeito de uma alteração unitária de  $X$  sobre o valor médio da variável  $Y$ , quando todas as outras variáveis estão fixas)

$E$  Erro aleatório ou estocástico (reflete a influência de outros factores no no comportamento da variável  $Y$  que não podem ser explicadas linearmente pelo comportamento da variável  $X$ )



**Resíduo (e) : Diferença entre o valor previsto (i.e., o ponto na linha) e o valor observado**

- **Modelo de regressão linear (simples)**
  - A utilização do OLS como método de estimação vai produzir uma série de medidas que nos vão permitir avaliar o valor explicativo (ou qualidade do ajustamento) do nosso modelo
  - A OLS é uma técnica de otimização matemática que procura encontrar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (i.e., dos resíduos)
  - A utilização do OLS como método de estimação vai produzir uma série de medidas que nos vão permitir avaliar o valor explicativo (ou qualidade do ajustamento) do nosso modelo



- **Modelo de regressão linear (simples)**

- **Variação Total de Y à volta da sua média (SST)**

=

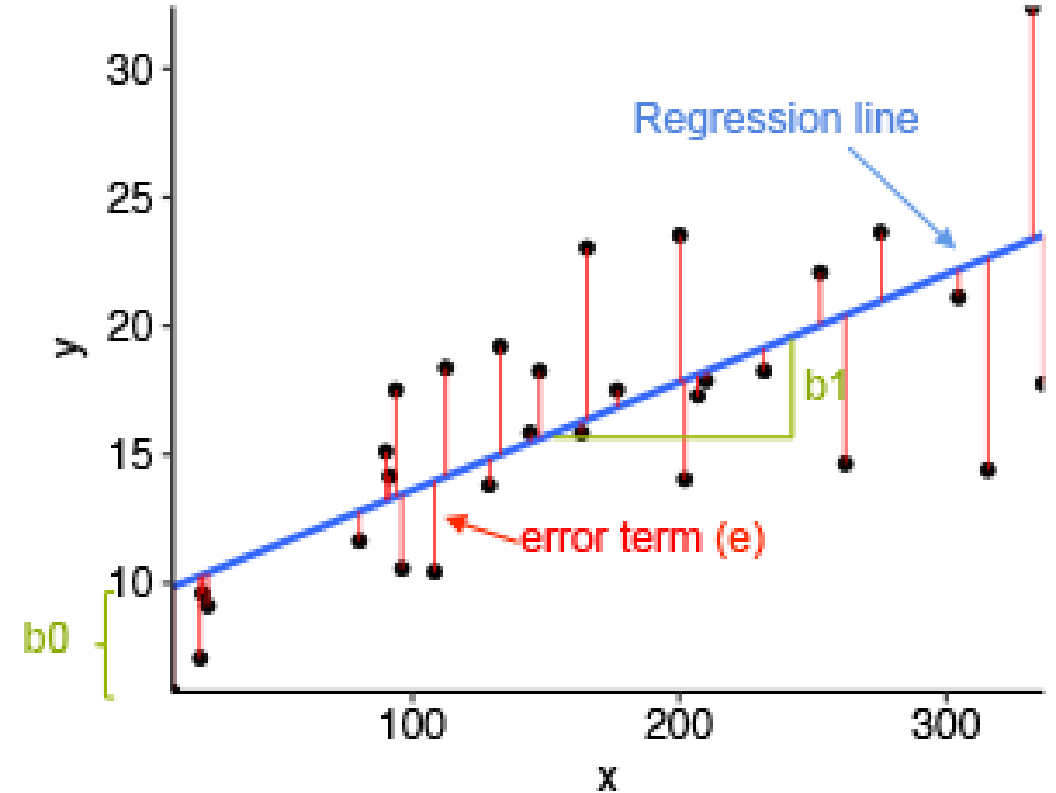
Varição explicada pelo modelo (SSR) = Soma dos Quadrados da Regressão

+

Varição não explicada (SSE) = Soma dos Quadrados dos Resíduos

- **A partir destas medidas podemos calcular o Coeficiente de Determinação (R<sup>2</sup>) que mede a proporção da Variação Total que é explicada pelo modelo:**

$$R^2 = SSR / SST$$





# ANÁLISE DE DADOS EM GRH

## *Aula 9: O modelo de regressão linear*

**Vamos então ver como isto funciona na prática**

# ANÁLISE DE DADOS EM GRH

## Aula 9: O modelo de regressão linear

- Quando produzimos um modelo de regressão no SPSS, o programa irá produzir 4 tabelas nos dão um conjunto de informações que devemos ter em conta.
- A primeira é a Tabela ‘Variáveis Inseridas/Removidas’
  - A** Esta coluna indica as variáveis incluídas no modelo.
  - B** Esta coluna indica-nos que variáveis foram excluídas do modelo. Com o método ‘inserir’ esta coluna permanecerá vazia.
  - C** Esta coluna indica o método que o SPSS usou para executar a regressão.

Variáveis Inseridas/Removidas<sup>a</sup>

Modelo	Variáveis inseridas <b>A</b>	Variáveis removidas <b>B</b>	Método <b>C</b>
1	Age, Height <sup>b</sup>	.	Inserir

a. Variável Dependente: Weight

b. Todas as variáveis solicitadas inseridas.

Fonte: <http://soniavieira.blogspot.com/2019/10/regressao-linear-multipla-no-spss.html>



# ANÁLISE DE DADOS EM GRH

## Aula 9: O modelo de regressão linear

- A Tabela 'ANOVA' dá-nos os resultados de um teste sobre a significância estatística do nosso modelo de regressão:

Os valores  $Z^A$  e  $Sig.^b$  mostram os resultado do teste F à hipótese nula de que:

Todas variáveis independentes do nosso modelo têm um coeficiente  $\beta$  igual a 0 (ou seja, o nosso modelo não é estatisticamente significativo).

Um valor de  $Z^A$  acima de 0 significa que haverá pelo menos uma variável no modelo que apresenta um coeficiente  $\beta$  diferente de 0.

Para que o modelo seja significativo o valor  $Sig.^b$  tem de ser menor que 0.05.

ANOVA<sup>a</sup>

Modelo	Soma dos Quadrados	df	Quadrado Médio	Z <sup>A</sup>	Sig. <sup>B</sup>
1					
Regressão	692,823	2	346,411	15,953	,001 <sup>b</sup>
Resíduo	195,427	9	21,714		
Total	888,250	11			

a. Variável Dependente: Weight

b. Preditores: (Constante), Age, Height

Fonte: <http://soniavieira.blogspot.com/2019/10/regressao-linear-multipla-no-spss.html>

# ANÁLISE DE DADOS EM GRH

## Aula 9: O modelo de regressão linear

- Depois temos a Tabela ‘Resumo do modelo’

**A**  $R^2$  é a proporção da variância na variável dependente que pode ser prevista a partir das variáveis independentes.

*Ex: As variáveis no modelo explicam 78% da variância da variável dependente.*

**B**  $R^2$  ajustado, é uma medida alternativa da capacidade explicativa do modelo. É mais adequado quando o número de observações é pequeno e o número de variáveis independentes é grande.

Este tipo de indicador vai ser muito importante quando compararmos diferentes modelos de regressão!

(mas o AIC também é bastante útil!)

Resumo do modelo

Modelo	R	R quadrado <b>A</b>	R quadrado ajustado <b>B</b>	Erro padrão da estimativa
1	,883 <sup>a</sup>	,780	,731	4,65984

a. Preditores: (Constante), Age, Height

Fonte: <http://soniavieira.blogspot.com/2019/10/regressao-linear-multipla-no-spss.html>

- A Tabela ‘Coeficientes’

**A** Os ‘Coeficientes não padronizados’ indicam duas coisas:

- O valor da ‘Constante’ na nossa tabela (**B**);
- O valor dos ‘Coeficientes não- padronizados (**C**), i.e. o quanto a variável dependente varia em função do aumento em uma unidade da variável independente, quando as outras variáveis independentes são mantidas constantes.

Na realidade, estes são os coeficientes que entram na equação de regressão

$$Y = \overset{B}{\beta_0} + \overset{C}{\beta_1}X + E$$

Coeficientes<sup>a</sup>

Modelo		Coeficientes não padronizados		Erro Padrão	Coeficientes padronizados		t	Sig.
		<b>A</b>	B		<b>D</b>	Beta		
1	(Constante)	<b>B</b>	6,553	10,945			,599	,564
	Height		<b>C</b> ,722	,261	,548		2,768	,022
	Age		2,050	,937	,433		2,187	,056

a. Variável Dependente: Weight

Fonte: <http://soniavieira.blogspot.com/2019/10/regressao-linear-multipla-no-spss.html>

- A Tabela ‘Coeficientes’

O que é que nos dizem estes coeficientes (<sup>C</sup>)?

Indicam o quanto varia a variável dependente em face do aumento de 1 unidade na variável independente.

**Ex:** Por cada ano adicional de vida dos indivíduos da amostra, o seu peso aumenta em 2,050 (Kg).

Coeficientes<sup>a</sup>

Modelo	Coeficientes não padronizados		Erro Padrão	Coeficientes padronizados		t	Sig.
	A	B		D	Beta		
1	(Constante)	<b>B</b> 6,553	10,945			,599	,564
	Height	<b>C</b> ,722	,261		,548	2,768	,022
	Age	2,050	,937		,433	2,187	,056

a. Variável Dependente: Weight

Fonte: <http://soniavieira.blogspot.com/2019/10/regressao-linear-multipla-no-spss.html>

- A Tabela ‘Coeficientes’

- Os ‘Coeficientes não padronizados’ não podem ser comparados entre si para determinar qual deles é mais influente no modelo, pois podem ser medidos em diferentes escalas
- Essa informação é nos dada pelos ‘Coeficientes padronizados’ (D)
- Neste caso, podemos concluir que a influência da altura (‘Height’) sobre o peso dos indivíduos é superior à influência da idade (‘Age’)

Coeficientes<sup>a</sup>

Modelo	Coeficientes não padronizados			Coeficientes padronizados		t	Sig.
	A	B	Erro Padrão	D	Beta		
1 (Constante)	B	6,553	10,945			,599	,564
Height		C	,722	,261	,548	2,768	,022
Age			2,050	,937	,433	2,187	,056

a. Variável Dependente: Weight

Fonte: <http://soniavieira.blogspot.com/2019/10/regressao-linear-multipla-no-spss.html>

- A Tabela ‘Coeficientes’

Como é que sabemos que o efeito de uma variável independente é estatisticamente significativo?

Olhando para o valor de Sig. :

A interpretação deste valor depende do grau de confiança que queremos adoptar

Grau de confiança = 95% → Sig. tem de ser  $\leq 0.05$

Grau de confiança = 99% → Sig. tem de ser  $\leq 0.01$

Grau de confiança = 90% → Sig. tem de ser  $\leq 0.1$

Neste caso, a variável ‘Altura’ tem uma relação estatisticamente significativa, para um grau de confiança de 95%

Coeficientes<sup>a</sup>

Modelo	Coeficientes não padronizados		Erro Padrão	Coeficientes padronizados		t	Sig.
	A	B		D	Beta		
1	(Constante)	B	6,553	10,945		,599	,564
	Height	C	,722	,261	,548	2,768	,022
	Age		2,050	,937	,433	2,187	,056

a. Variável Dependente: Weight

Fonte: <http://soniavieira.blogspot.com/2019/10/regressao-linear-multipla-no-spss.html>

## Implementação do Modelo de Regressão Linear

- 1. Inspeccionar a variável independente*
- 2. Explorar as relações entre variáveis*
- 3. Definir um modelo de regressão linear (IV contínuas)*
- 4. Adicionar IVs nominais*
- 5. Qual é o modelo mais eficiente? (I)*
- 6. Qual é o modelo mais eficiente (II)? A Regressão Stepwise*

## Modelo de Regressão Linear

- **Objectivo:**
  - **Avaliar em que medida a distribuição dos salários na empresa variam em função da experiência dos trabalhadores e das avaliações que os seus superIVsors**



# Implementação do Modelo de Regressão Linear

## *1. Inspeccionar a variável independente*

*Histograma*

*Gráfico de Caixas*

*Gráfico de Q*



# Implementação do Modelo de Regressão Linear

- 1. Inspeccionar a variável independente*  
*Histograma*

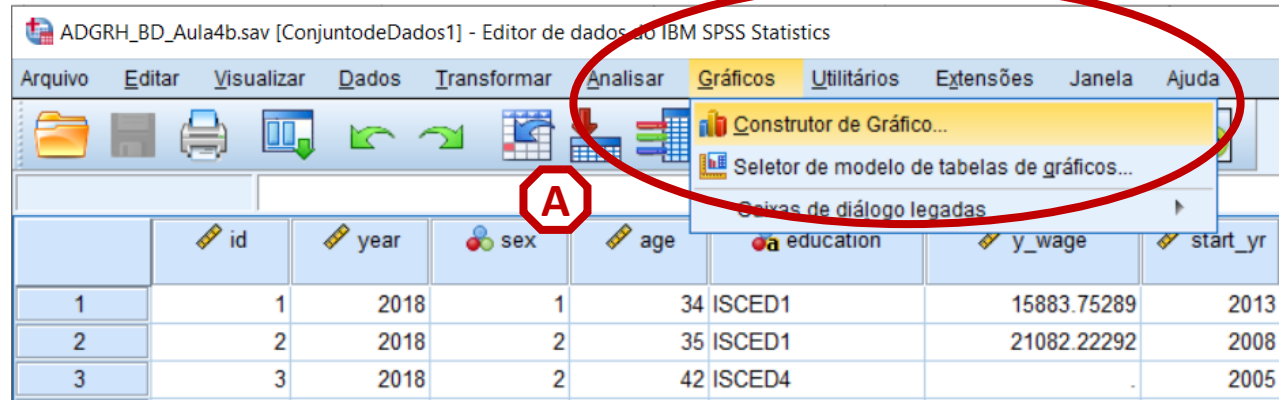
## Modelo de Regressão Linear

- Objectivo:
  - Avaliar como se distribuem os salários na organização

## Histograma

- Selecionar 'Gráficos' / 'Construtor de Gráfico'

A



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

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

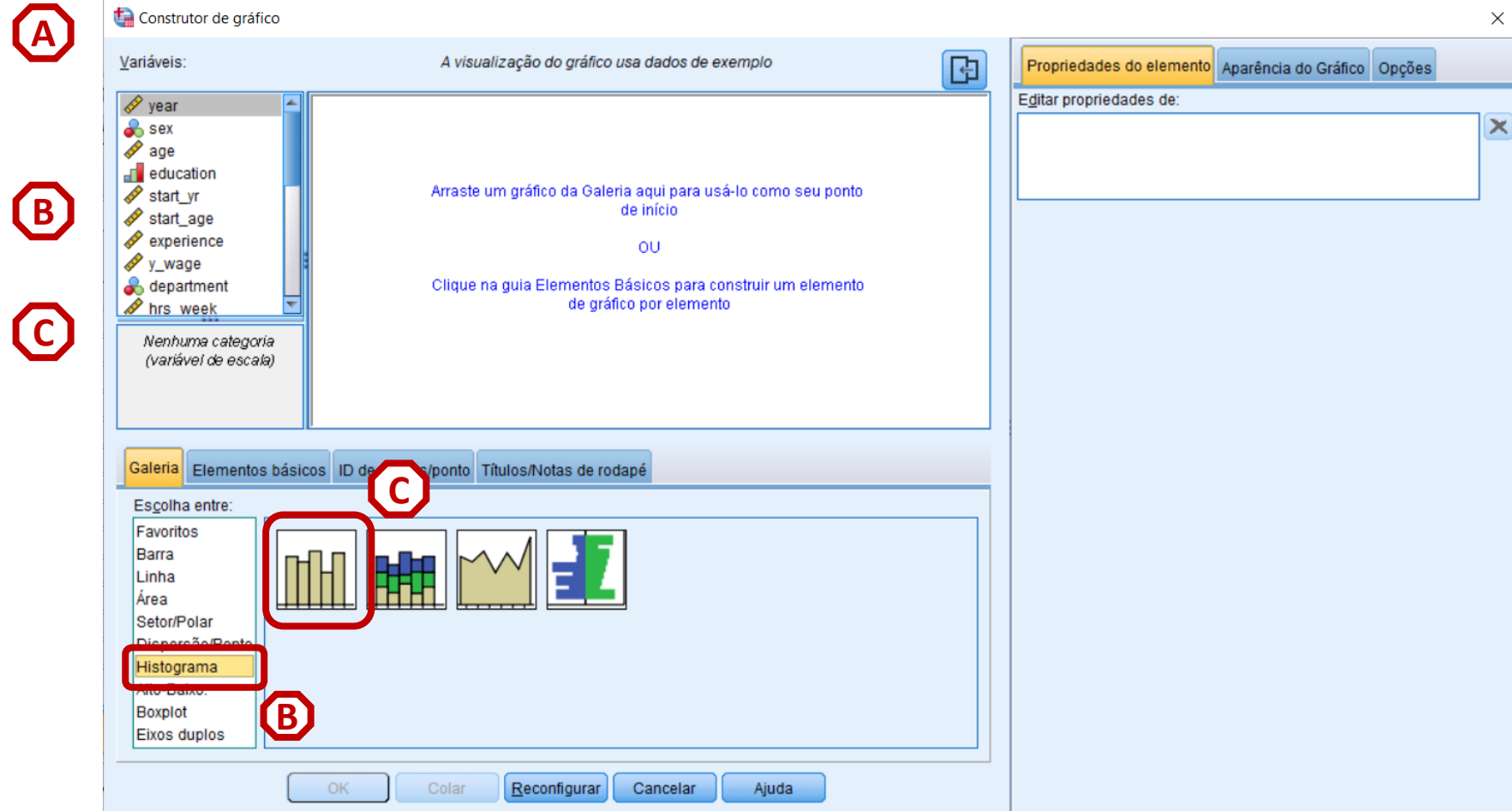
Construtor de Gráfico...  
Seletores de modelo de tabelas de gráficos...

Caixas de diálogo legadas

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

## Histograma

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'Histograma'
- Duplo-clique no Histograma Simples



**A**

**B**

**C**

**B**

## Histograma

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'Histograma'
- Duplo-clique no Histograma Simples
- Seleccionar a variável 'y\_wage2'
- E arrastar para o 'Eixo X'

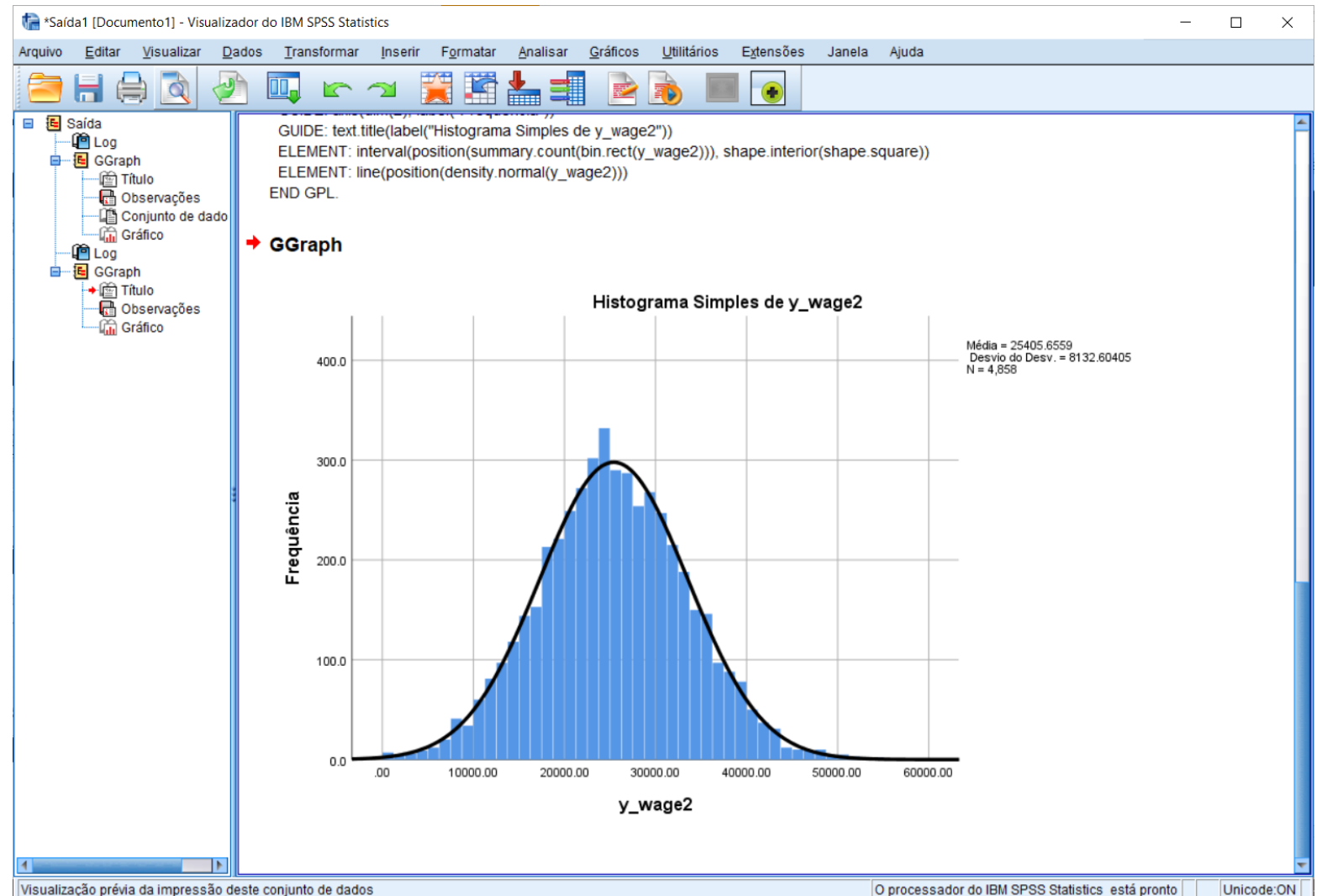
The screenshot shows the 'Construtor de gráfico' (Chart Builder) interface. The window title is 'Construtor de gráfico'. The main area displays a 'Histograma Simples' chart. The chart area has a red box around the 'Eixo X?' label and a red circle 'E' next to it. The 'Variáveis:' list on the left includes 'muscle\_lower', 'headache', 'injury', 'anxiety', 'fatigue', 'absent\_nr', 'workcond\_sat', 'evaluation', 'y\_wage2', and 'evaluation2'. The 'y\_wage2' variable is highlighted with a red circle 'D'. The 'Galeria' at the bottom shows various chart types, with 'Histograma' selected. The right panel shows the 'Propriedades do elemento' for 'Barra1', with 'Estadística' set to 'Histograma'. The 'Eixo X?' label is highlighted with a red box, and a red circle 'E' is next to it. A red circle 'A' is on the top left, 'B' on the top left of the chart area, and 'C' on the top left of the chart area.

### Histograma

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'Histograma'
- Duplo-clique no Histograma Simples
- Seleccionar a variável 'y\_wage2'
- E arrastar para o 'Eixo X'
- Seleccionar 'Exibir curva Normal'
- Seleccionar 'OK'

## Histograma

- O gráfico é publicado no 'Visualizador de Resultados'
- Não se encontram grandes problemas na variável...







# Implementação do Modelo de Regressão Linear

## *1. Inspeccionar a variável independente*

*Gráfico de Caixa*

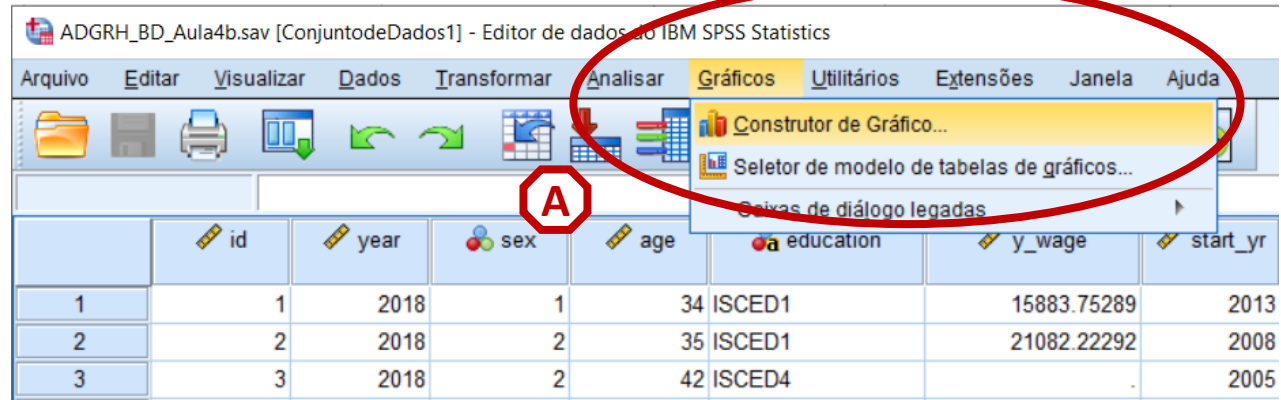
## Modelo de Regressão Linear

- **Objectivo:**
  - **Detectar a existência de casos extremos e remediar a situação**

### Gráfico de Caixa

- Selecionar 'Gráficos' / 'Construtor de Gráfico'

A



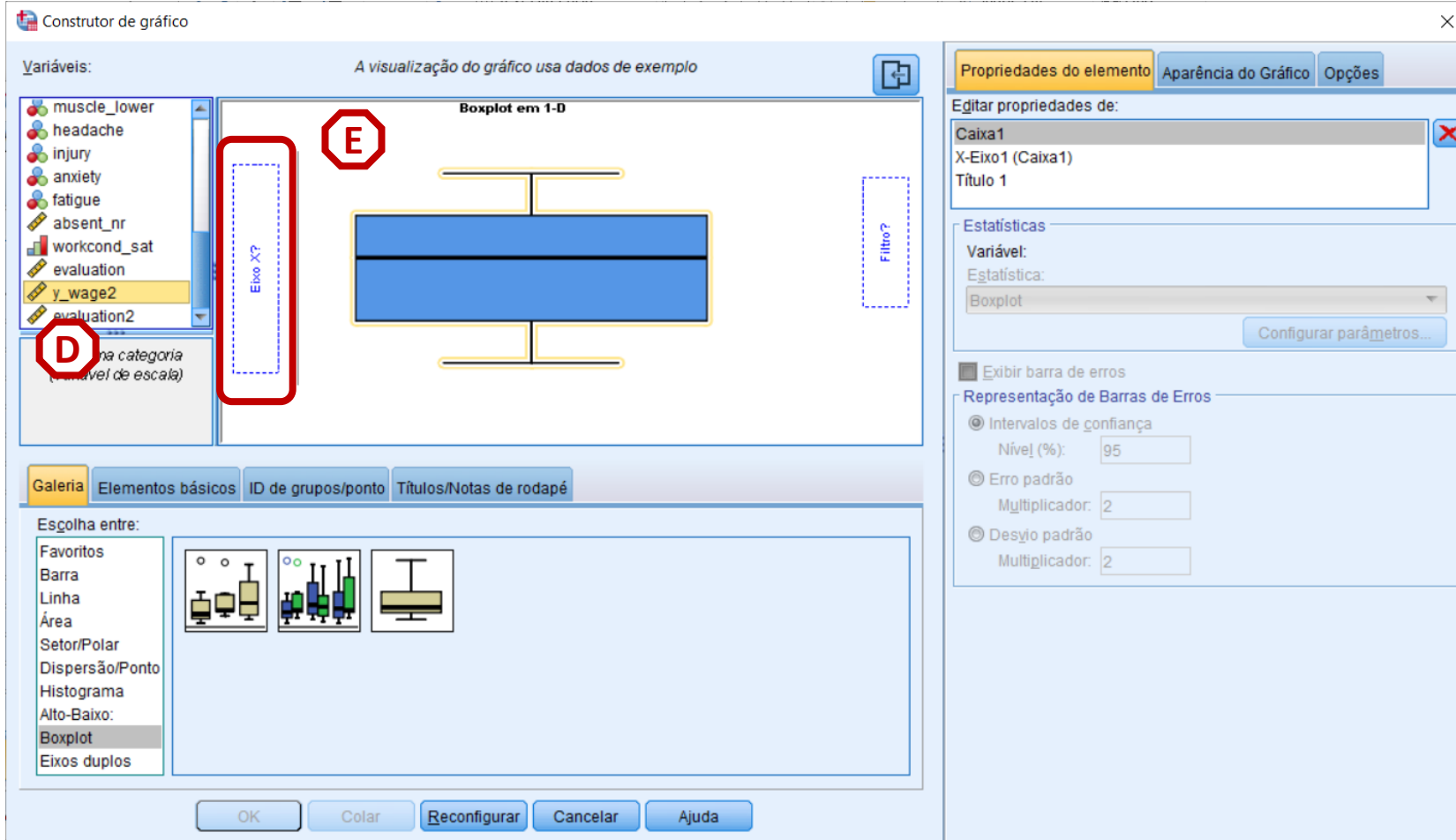
### Gráfico de Caixa

- 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'. A dashed box labeled 'A' is around this list. In the center, a 'Boxplot em 1-D' is displayed. On the right, a properties panel is open for 'Caixa1', showing options for 'Estatística' (set to 'Boxplot') and 'Exibir barra de erros'. Below the main window, a 'Galeria' (Gallery) shows various chart types. A dashed box labeled 'B' is around the 'Boxplot' option, and another dashed box labeled 'C' is around the simple boxplot icon. 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\_wage2'
- E arrastar para o 'Eixo X'



**Construtor de gráfico**

A visualização do gráfico usa dados de exemplo

**Variáveis:**

- muscle\_lower
- headache
- injury
- anxiety
- fatigue
- absent\_nr
- workcond\_sat
- evaluation
- y\_wage2**
- evaluation2

**Boxplot em 1-D**

Eixo X?

Filtro?

**Propriedades do elemento** Aparência do Gráfico Opções

Editar propriedades de:

Caixa1

X-Eixo1 (Caixa1)

Título 1

**Estatísticas**

Variável:

Estatística:

Boxplot

Configurar parâmetros...

Exibir barra de erros

**Representação de Barras de Erros**

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

OK Colar Reconfigurar Cancelar Ajuda

### Gráfico de Caixa

- Seleccionar 'Gráficos' / 'Construtor de Gráfico'
- Seleccionar 'Boxplot'
- Duplo-clique no Gráfico de Caixa (simples)
- Seleccionar a variável 'y\_wage2'
- E arrastar para o 'Eixo X'
- Seleccionar 'OK'

A

B

D

C

D

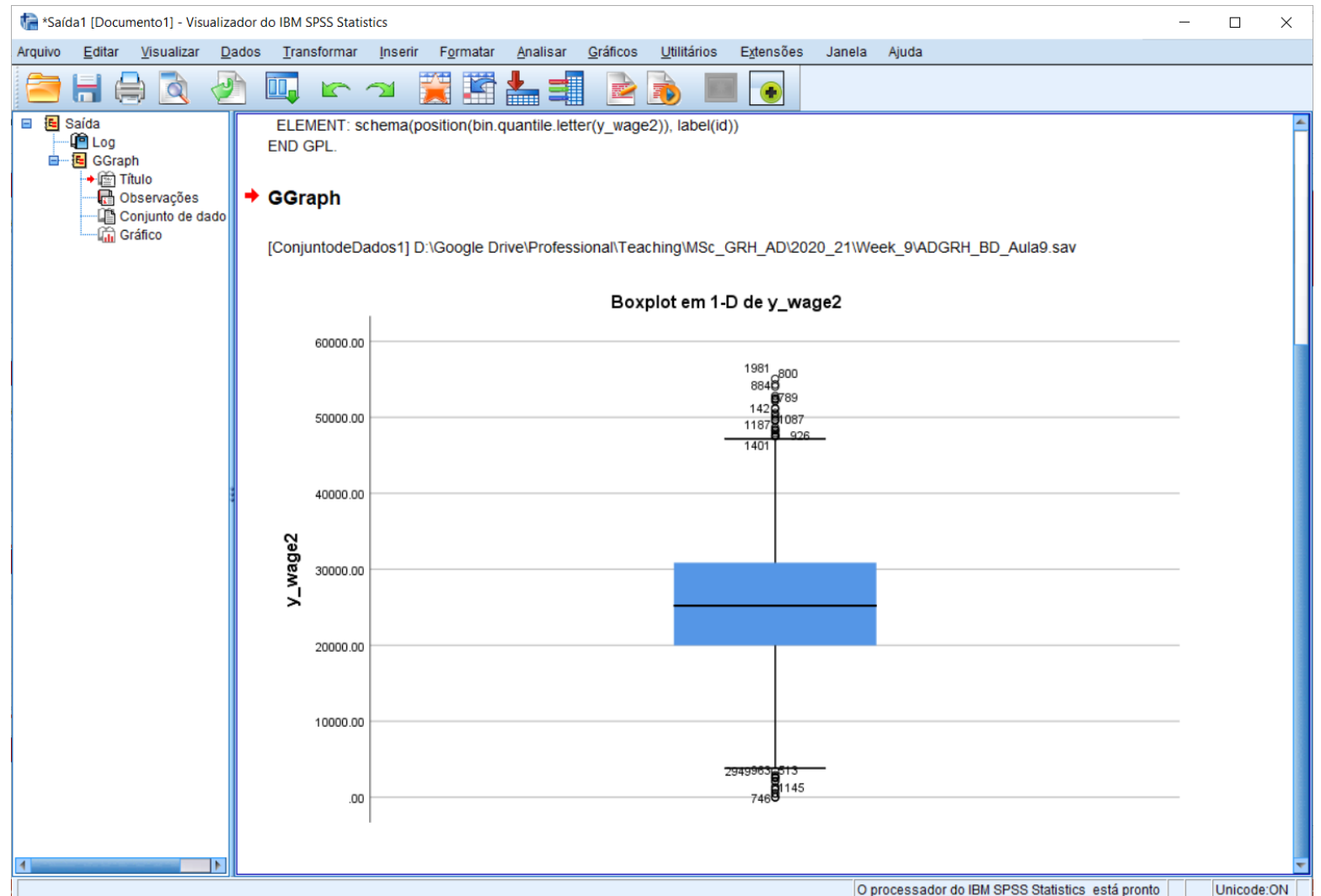
E

F

F

## Gráfico de Caixa

- O gráfico é publicado no 'Visualizador de Resultados'
- Sobre a forma como podemos lidar com estes casos extremos – ver a Aula 5!



# Implementação do Modelo de Regressão Linear

## *1. Inspeccionar a variável independente*

*Teste de Shapiro-Wilk & Gráfico de Q*



## Modelo de Regressão Linear

- Objectivo:
  - Avaliar se a variável segue uma distribuição normal

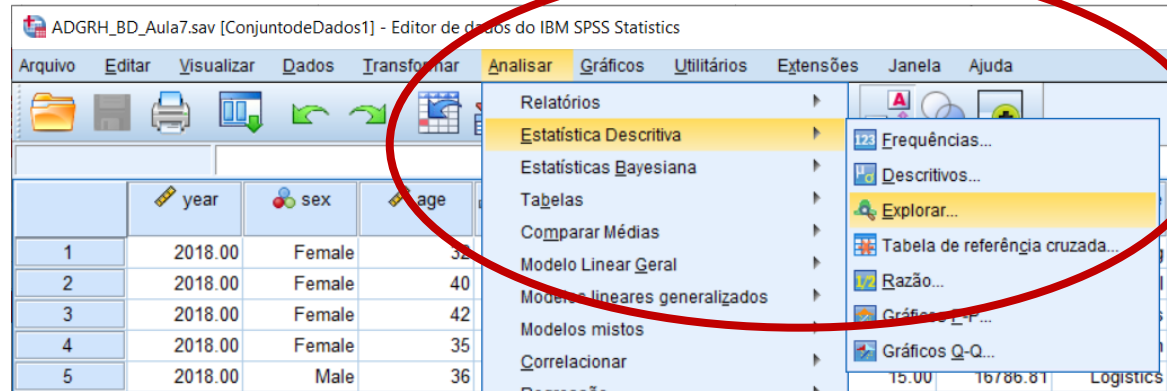
### Teste de Shapiro-Wilk

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Explorar'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Lista de Variáveis Dependentes'

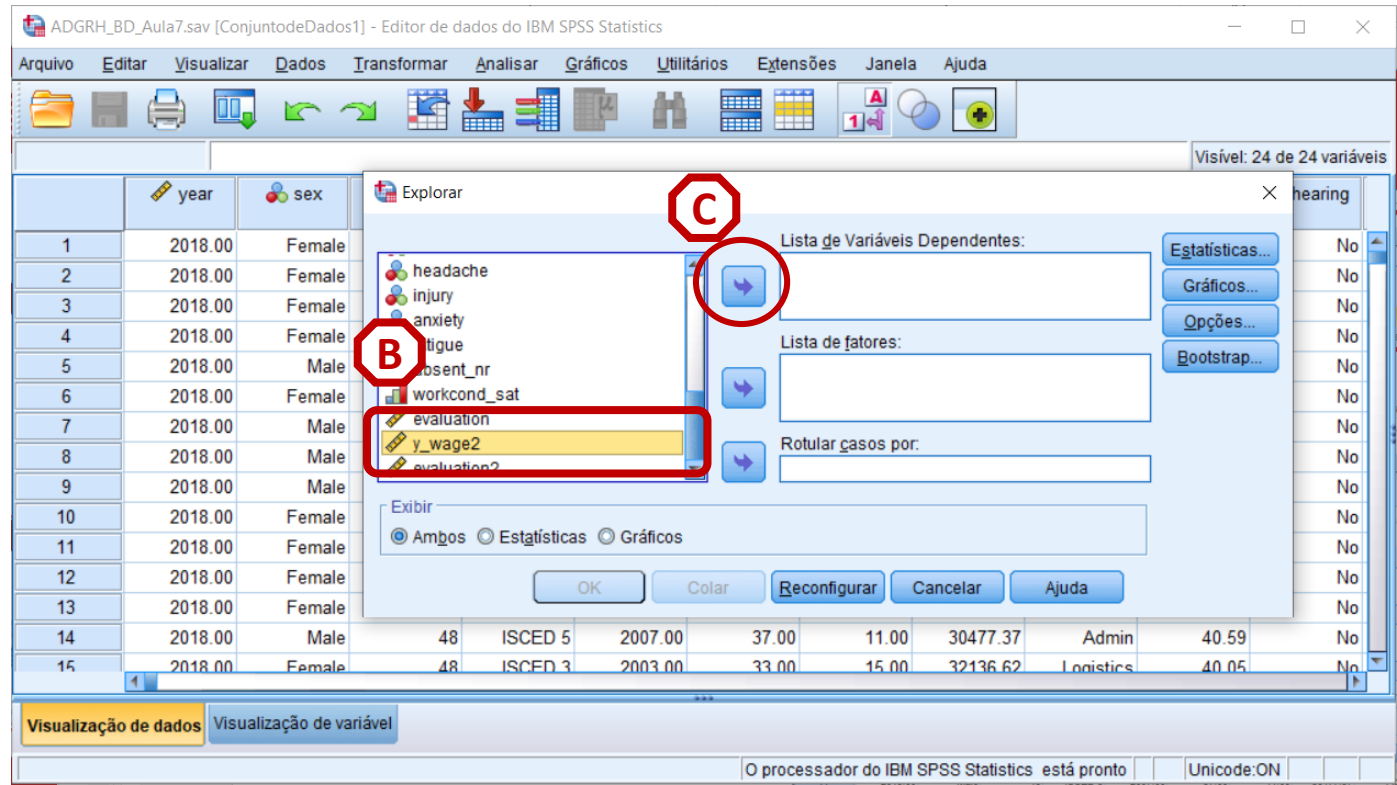
A

B

C



A



### Teste de Shapiro-Wilk

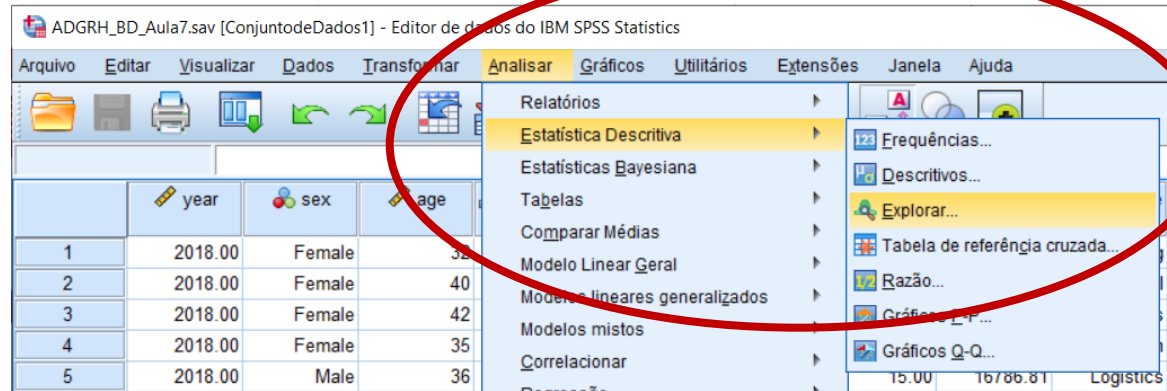
- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Explorar'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Lista de Variáveis Dependentes'
- Selecionar 'Gráficos'

**A**

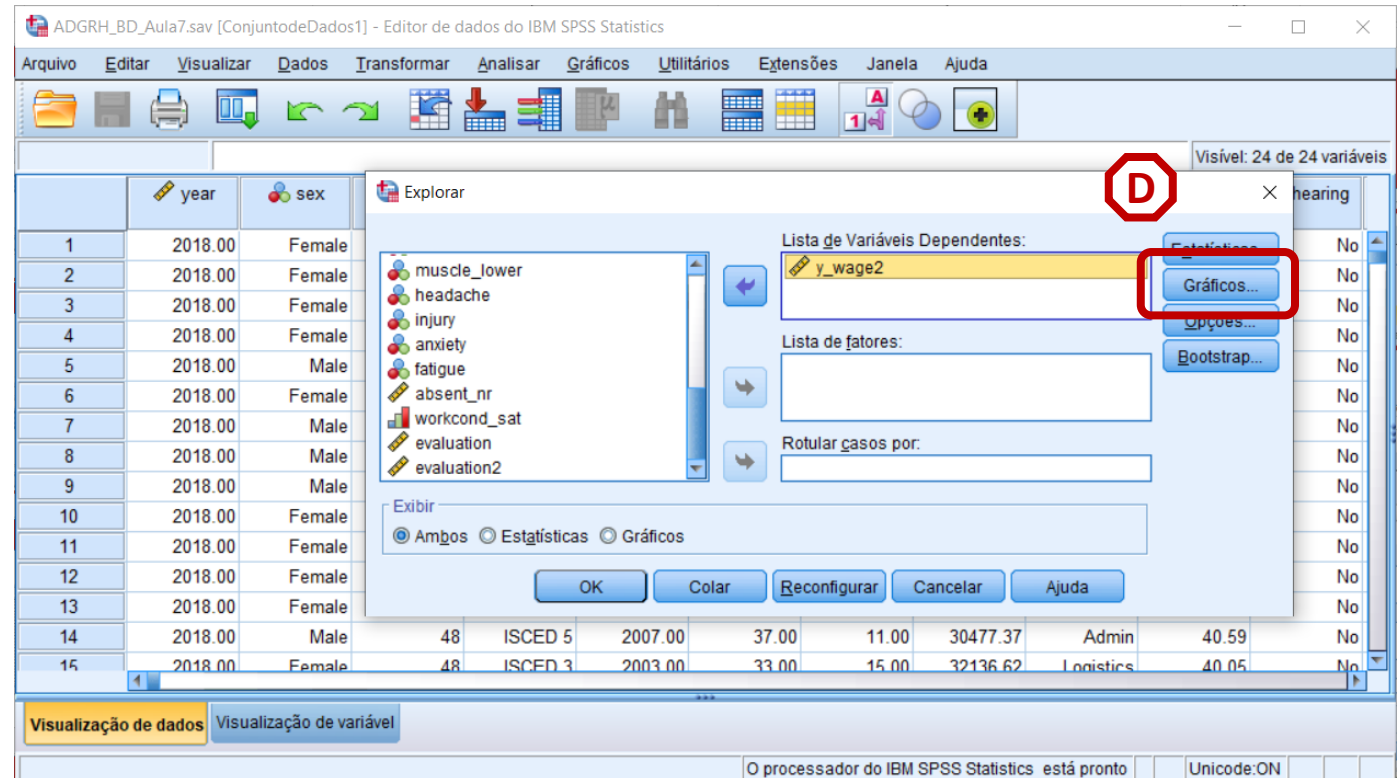
**B**

**C**

**D**



**A**



**D**

### Teste de Shapiro-Wilk

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Explorar'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Lista de Variáveis Dependentes'
- Selecionar 'Gráficos'
- Selecionar "Gráficos de normalidade com testes"
- Selecionar 'Continuar'

A

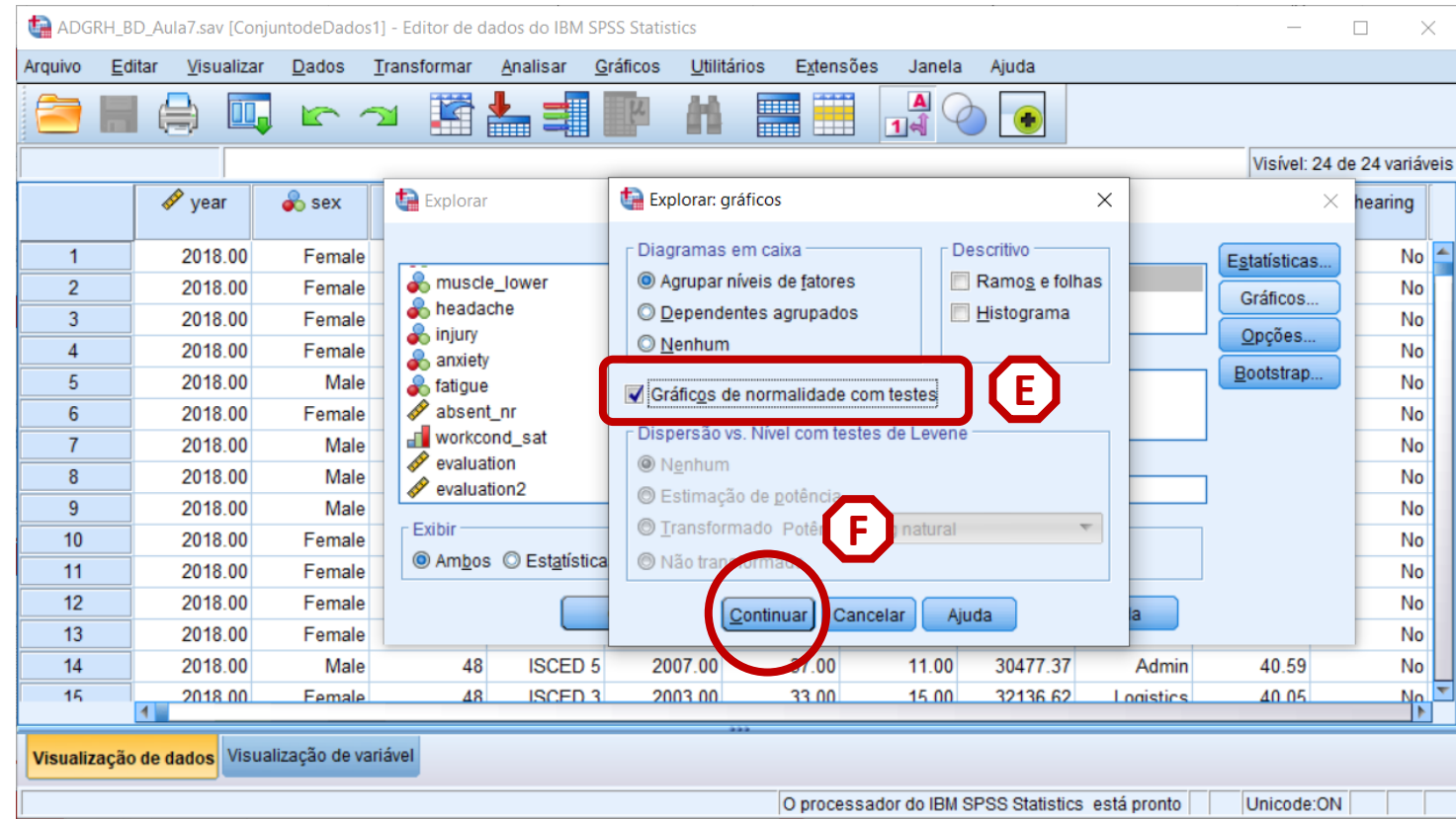
B

C

D

E

F



## Teste de Shapiro-Wilk

- Selecionar 'Opções'



ADGRH\_BD\_Aula7.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Visível: 24 de 24 variáveis

Explorar

Lista de Variáveis Dependentes:  
y\_wage2

Lista de fatores:

Rotular casos por:

Exibir  
 Ambos  Estatísticas  Gráficos

OK Colar Reconfigurar Cancelar Ajuda

	year	sex	muscle_lower	headache	injury	anxiety	fatigue	absent_nr	workcond_sat	evaluation	evaluation2	hearing
1	2018.00	Female										No
2	2018.00	Female										No
3	2018.00	Female										No
4	2018.00	Female										No
5	2018.00	Male										No
6	2018.00	Female										No
7	2018.00	Male										No
8	2018.00	Male										No
9	2018.00	Male										No
10	2018.00	Female										No
11	2018.00	Female										No
12	2018.00	Female										No
13	2018.00	Female										No
14	2018.00	Male	48	ISCED 5	2007.00	37.00	11.00	30477.37	Admin	40.59		No
15	2018.00	Female	48	ISCED 3	2003.00	33.00	15.00	32136.62	Logistics	40.05		No

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Teste de Shapiro-Wilk

- Selecionar 'Opções'
- Selecionar 'Excluir Casos por método pairwise'
- Selecionar 'Continuar'/OK



ADGRH\_BD\_Aula7.sav [Conjunto de Dados 1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Visível: 24 de 24 variáveis

Explorar

Lista de Variáveis Dependentes:

Excluir Casos

Excluir Casos por método pairwise

Opções

Valores omissos

Excluir casos pelo método listwise

Excluir casos por método pairwise

Relatar valores

Continuar Cancelar Ajuda

Exibir

Ambos Estatísticas Gráficos

OK Colar Reconfigurar Cancelar Ajuda

	year	sex	muscle_lower	headache	injury	anxiety	fatigue	absent_nr	workcond_sat	evaluation	evaluation2	hearing
1	2018.00	Female										No
2	2018.00	Female										No
3	2018.00	Female										No
4	2018.00	Female										No
5	2018.00	Male										No
6	2018.00	Female										No
7	2018.00	Male										No
8	2018.00	Male										No
9	2018.00	Male										No
10	2018.00	Female										No
11	2018.00	Female										No
12	2018.00	Female										No
13	2018.00	Female										No
14	2018.00	Male	48	ISCED 5	2007.00	37.00	11.00	30477.37	Admin	40.59		No
15	2018.00	Female	48	ISCED 3	2003.00	33.00	15.00	32136.62	Logistics	40.05		No

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

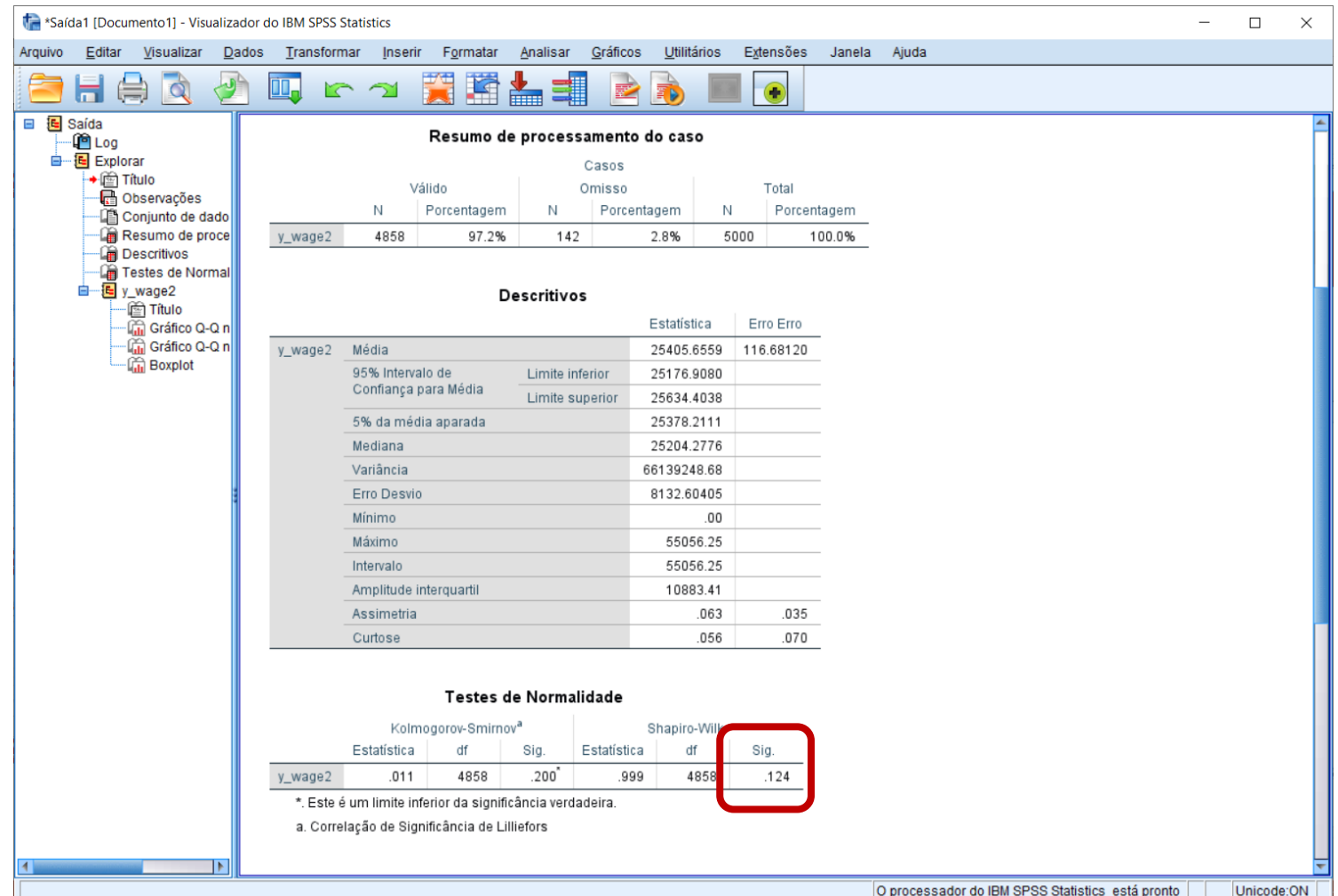
## Teste de Shapiro-Wilk

- O resultado é publicado no 'Visualizador de Resultados'

### INTERPRETAÇÃO:

- 'Sig'.  $\leq 0.05$ , rejeita-se a hipótese ( $H_0$ ) de que a variável segue uma distribuição normal. Aceita-se hipótese  $H_1$
- 'Sig'.  $> 0.05$ , não se rejeita a hipótese ( $H_0$ ) de que a variável segue uma distribuição normal.

A VARIÁVEL SALÁRIOS SEGUE UMA DISTRIBUIÇÃO NORMAL.



\*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

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

Resumo de processamento do caso

	Válido		Casos Omissos		Total	
	N	Porcentagem	N	Porcentagem	N	Porcentagem
y_wage2	4858	97.2%	142	2.8%	5000	100.0%

Descritivos

	Estatística	Erro
y_wage2 Média	25405.6559	116.68120
95% Intervalo de Confiança para Média	Limite inferior	25176.9080
	Limite superior	25634.4038
5% da média aparada	25378.2111	
Mediana	25204.2776	
Variância	66139248.68	
Erro Desvio	8132.60405	
Mínimo	.00	
Máximo	55056.25	
Intervalo	55056.25	
Amplitude interquartil	10883.41	
Assimetria	.063	.035
Curtose	.056	.070

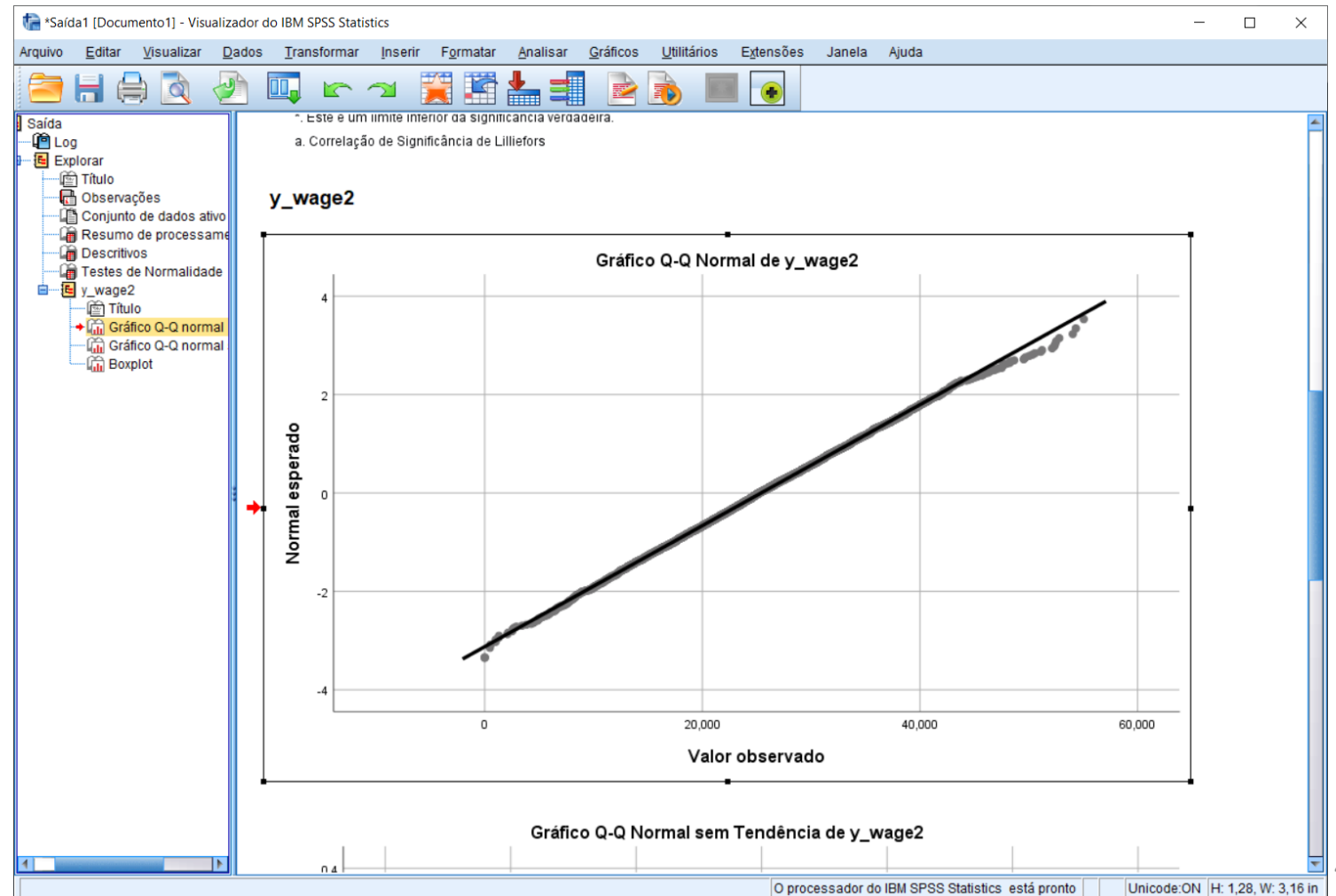
Testes de Normalidade

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estatística	df	Sig.	Estatística	df	Sig.
y_wage2	.011	4858	.200 <sup>*</sup>	.999	4858	.124

\*. Este é um limite inferior da significância verdadeira.  
a. Correlação de Significância de Lilliefors

## Teste de Shapiro-Wilk

- Confirma-se que a distribuição dos salários na empresa segue uma distribuição normal.





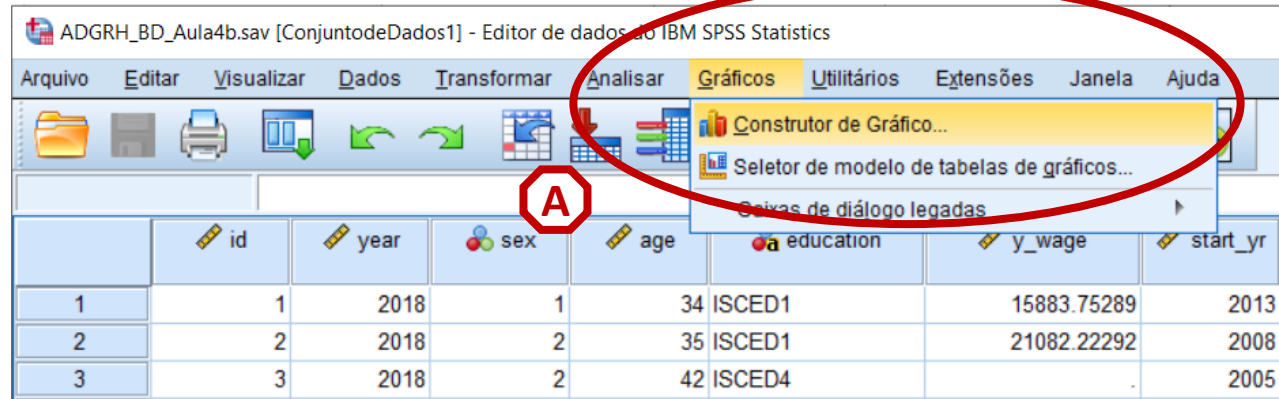


# Implementação do Modelo de Regressão Linear

*3. Explorar as relações entre variáveis*

### Matriz de Dispersão

- Seleccionar 'Gráficos' / 'Construtor de Gráfico'



## Matriz de Dispersão

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'Dispersão/Ponto'
- Duplo-clique em 'Matriz de Gráfico Disperso'

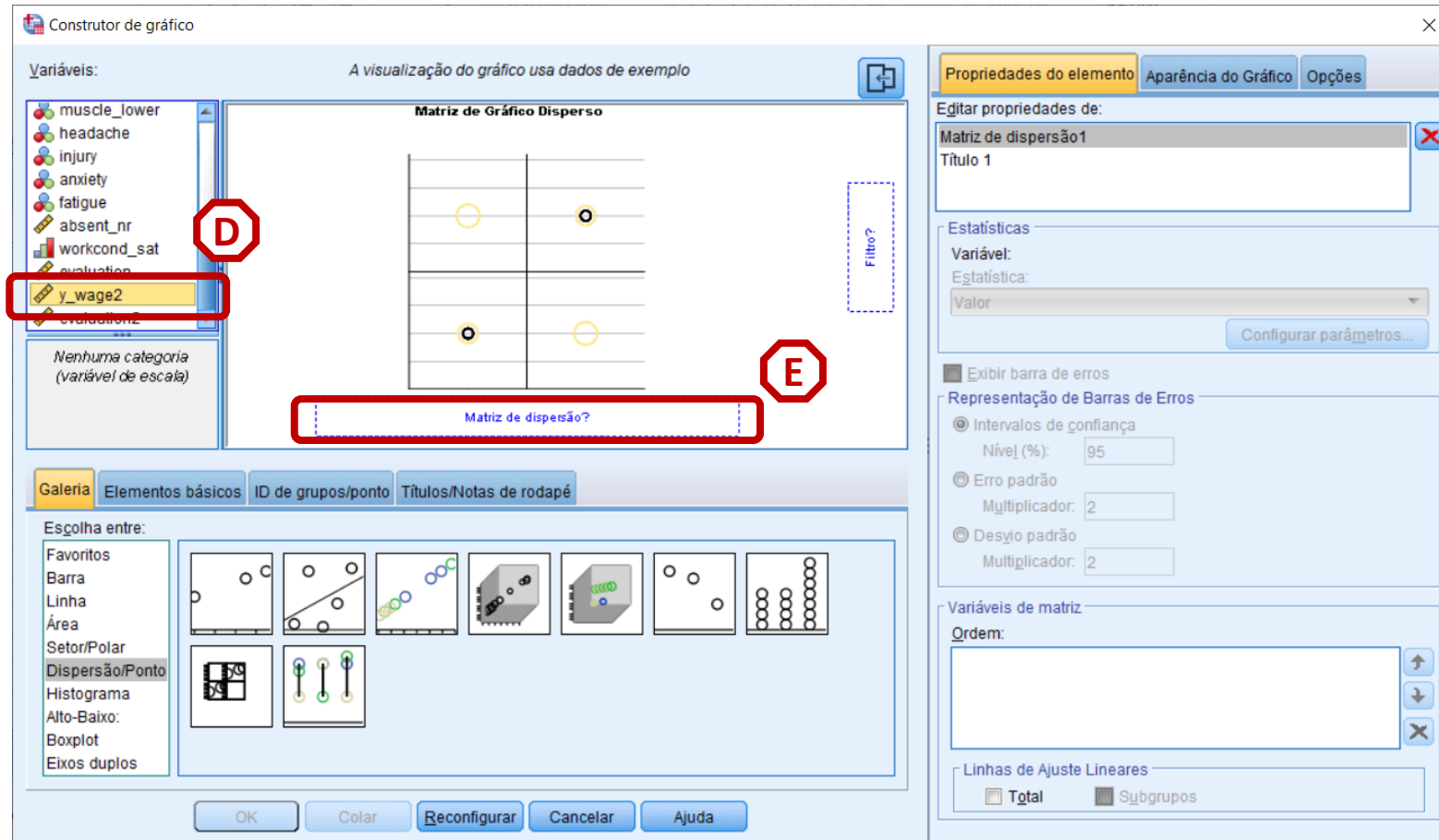
A

B

C

### Matriz de Dispersão

- Selecionar ‘Gráficos’ / ‘Construtor de Gráfico’
- Selecionar ‘Dispersão/Ponto’
- Duplo-clique em ‘Matriz de Gráfico Disperso’
- Selecionar a variável ‘y\_wage2’
- Arrastar para a caixa ‘Matriz de Dispersão’



**A** Construtor de gráfico

**B** Variáveis:

- muscle\_lower
- headache
- injury
- anxiety
- fatigue
- absent\_nr
- workcond\_sat
- evaluation
- y\_wage2**
- evaluation2

**C** Matriz de Gráfico Disperso

**D** Nenhuma categoria (variável de escala)

**E** Matriz de dispersão?

**E** Estatísticas

Variável:

Estatística:

Valor

Configurar parâmetros...

Exibir barra de erros

Representação de Barras de Erros

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

Variáveis de matriz

Ordem:

Linhas de Ajuste Lineares

- Total
- Subgrupos

OK Colar Reconfigurar Cancelar Ajuda

Exercício: Colocar as IVs ('experience', 'evaluation') na 'Matriz de Dispersão'

### Matriz de Dispersão

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'Dispersão/Ponto'
- Duplo-clique em 'Matriz de Gráfico Disperso'
- Selecionar a variável 'y\_wage2'
- Arrastar para a caixa 'Matriz de Dispersão'
- Clicar 'OK'

A

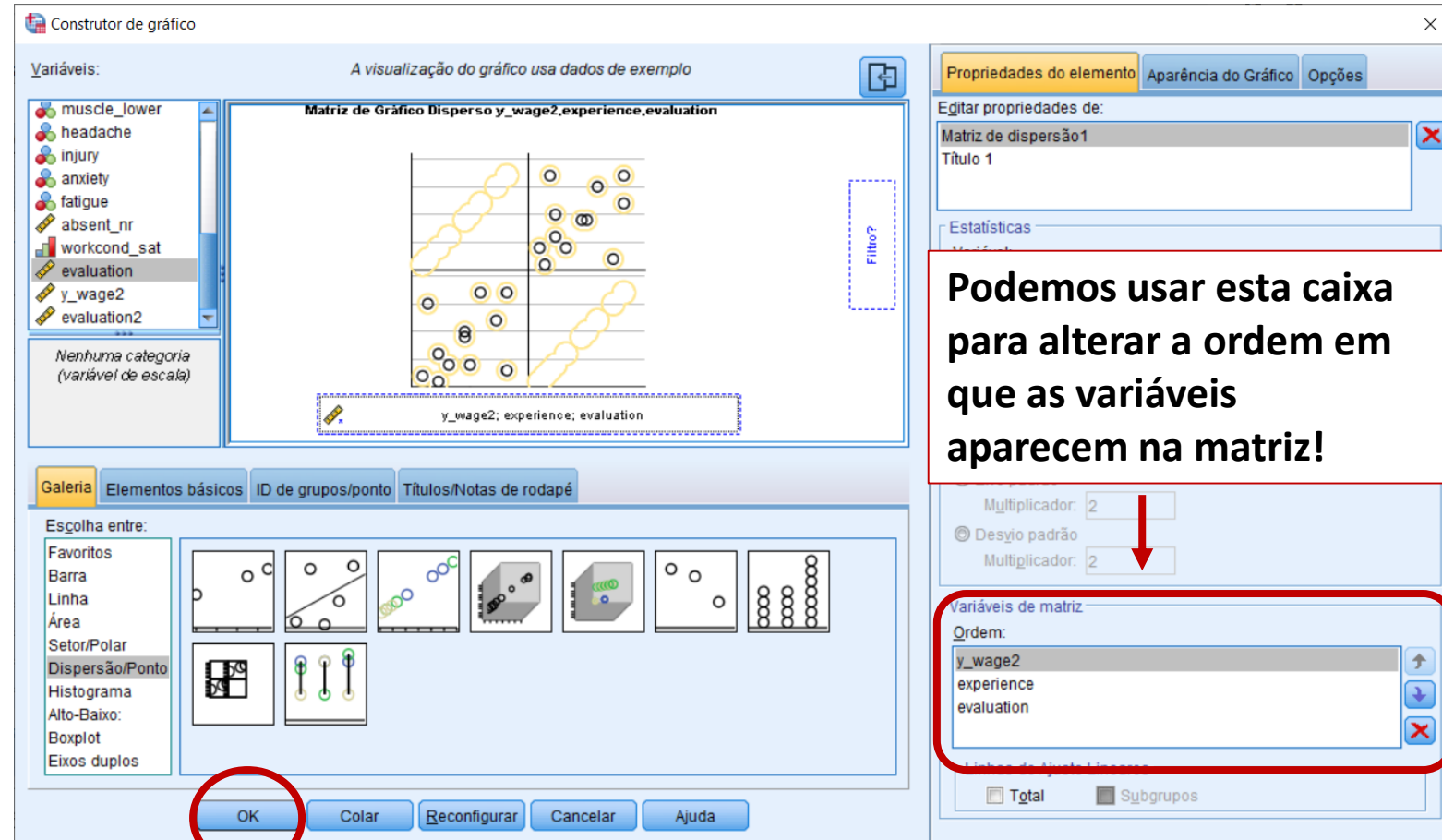
B

C

D

E

F



Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- muscle\_lower
- headache
- injury
- anxiety
- fatigue
- absent\_nr
- workcond\_sat
- evaluation
- y\_wage2
- evaluation2

Nenhuma categoria (variável de escala)

Matriz de Gráfico Disperso y\_wage2,experience,evaluation

Filtro?

y\_wage2; experience; evaluation

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:

Matriz de dispersão1

Título 1

Estatísticas

Podemos usar esta caixa para alterar a ordem em que as variáveis aparecem na matriz!

Multiplicador: 2

Desvio padrão

Multiplicador: 2

Variáveis de matriz

Ordem:

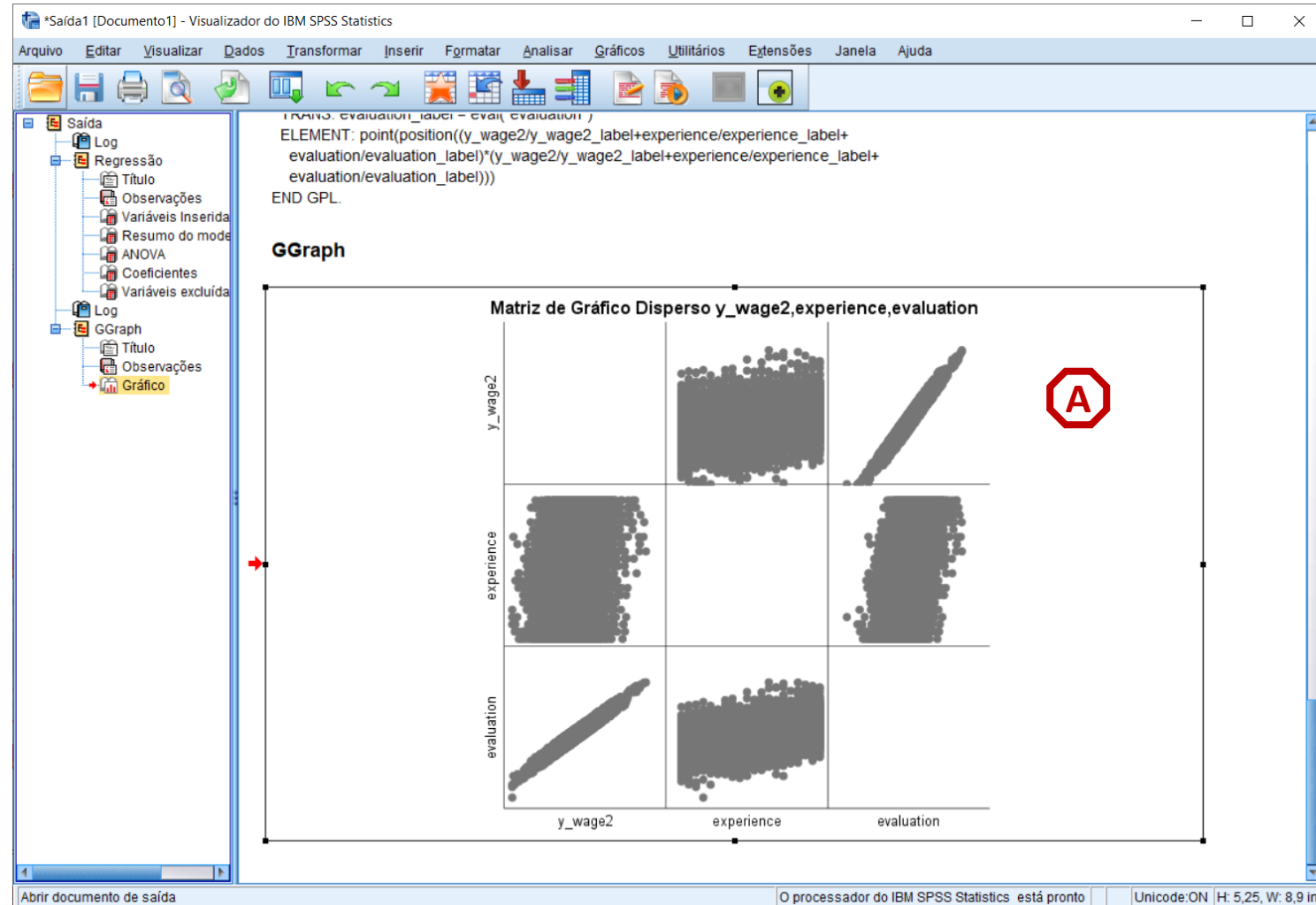
- y\_wage2
- experience
- evaluation

Total | Subgrupos

F

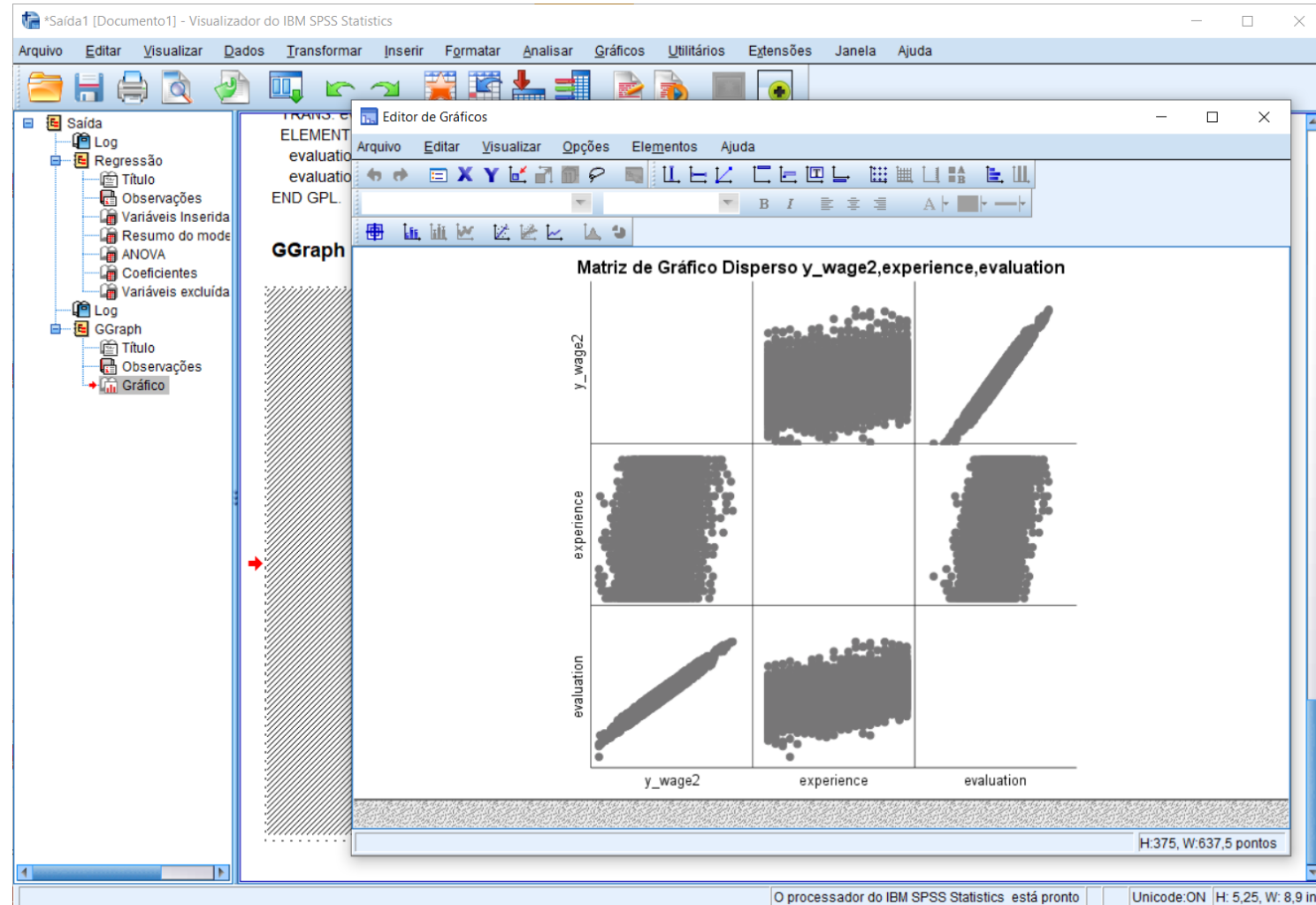
## Matriz de Dispersão

- O resultado é publicado no 'Visualizador de Resultados'
- Mas podemos melhorar a informação do gráfico
- Vamos fazer um duplo-clique sobre o gráfico



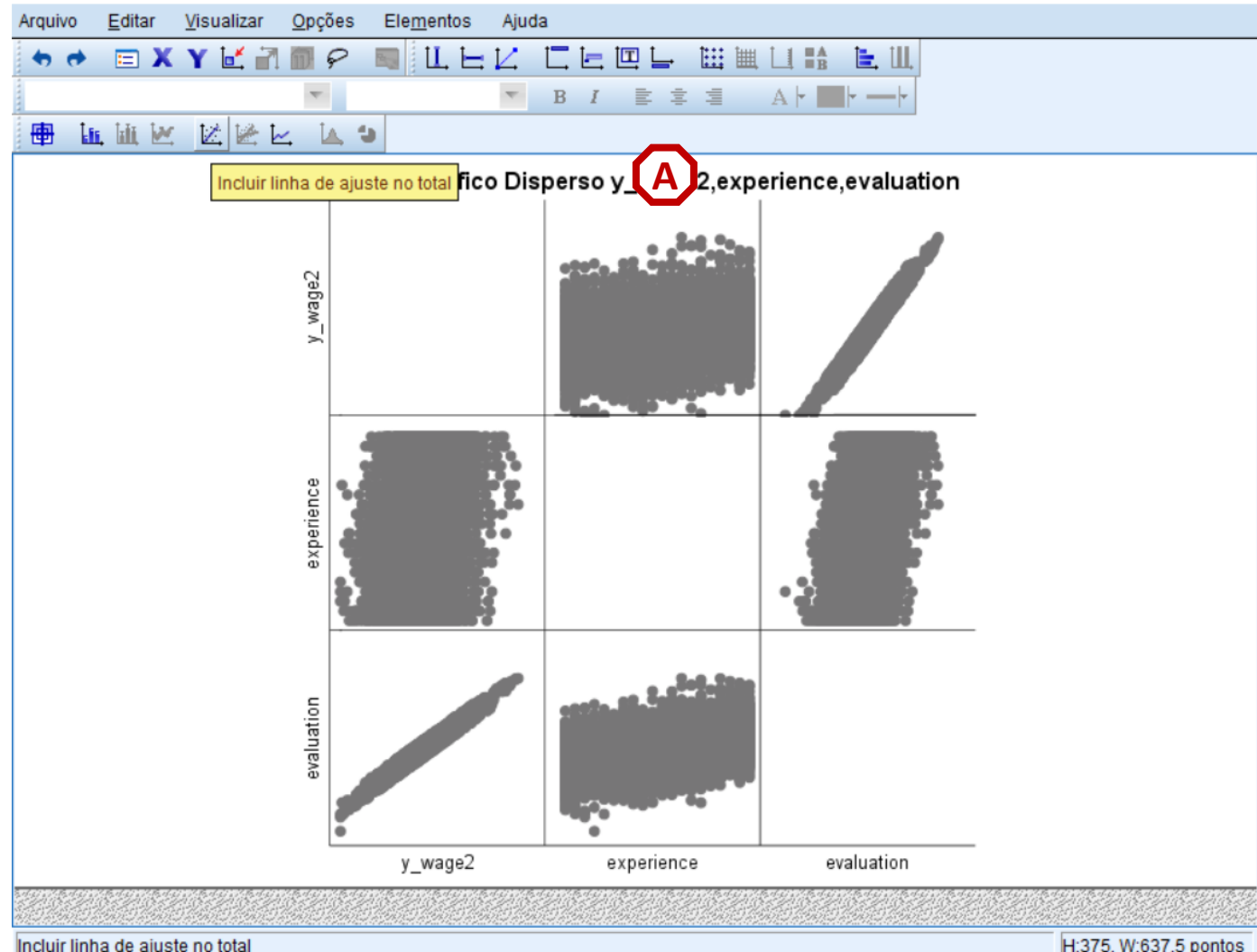
## Matriz de Dispersão

- O resultado é publicado no 'Visualizador de Resultados'
- Mas podemos melhorar a informação do gráfico
- Vamos fazer um duplo-clique sobre o gráfico
- Isso vai permitir abrir o 'Editor de Gráficos'



## Matriz de Dispersão

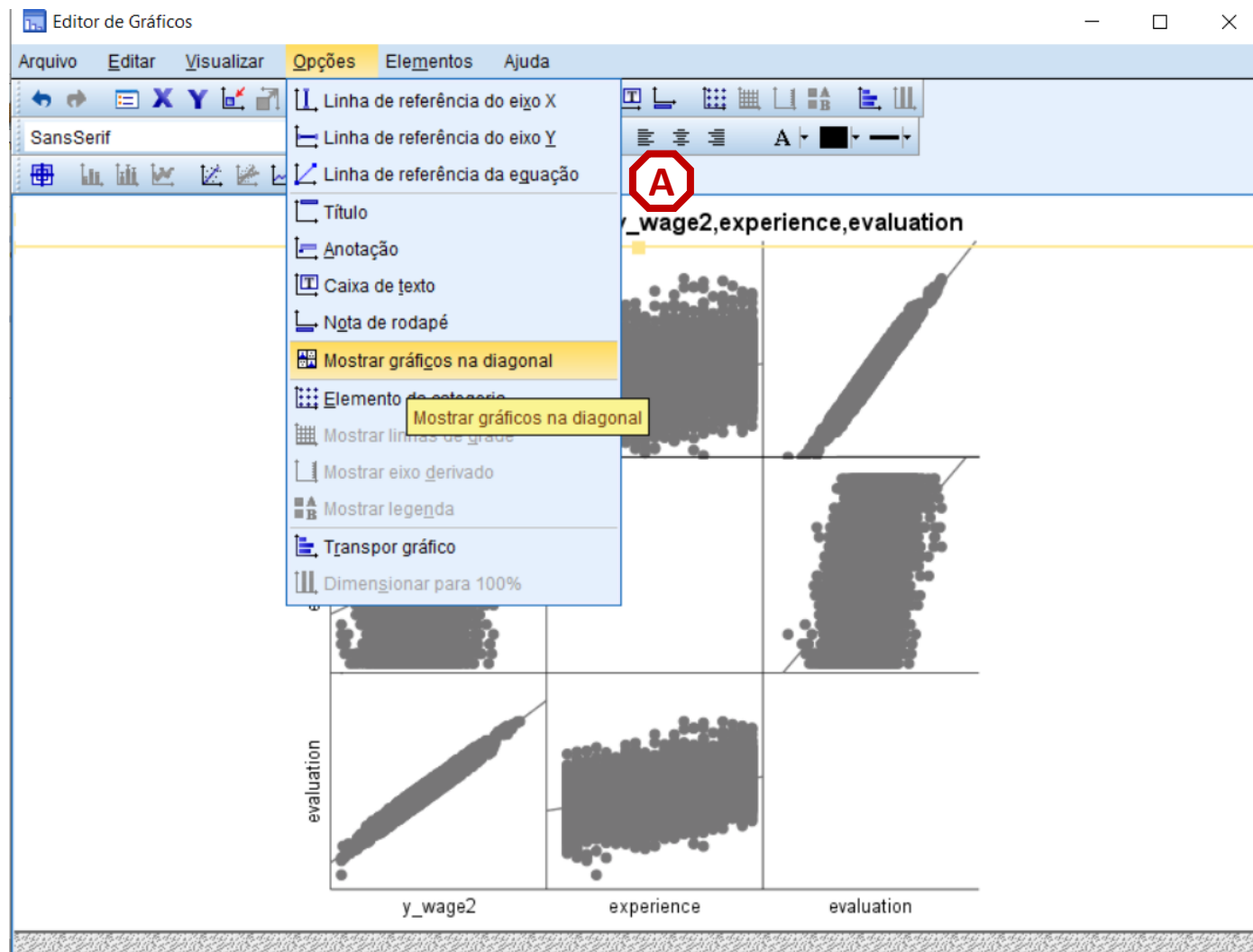
- No Editor de Gráficos...
- Vamos selecionar a opção 'Incluir linha de ajuste no total'





## Matriz de Dispersão

- No Editor de Gráficos...
- Vamos selecionar a opção 'Incluir linha de ajuste no total'
- Vamos querer também selecionar 'Opções' / 'Mostrar gráficos na diagonal'



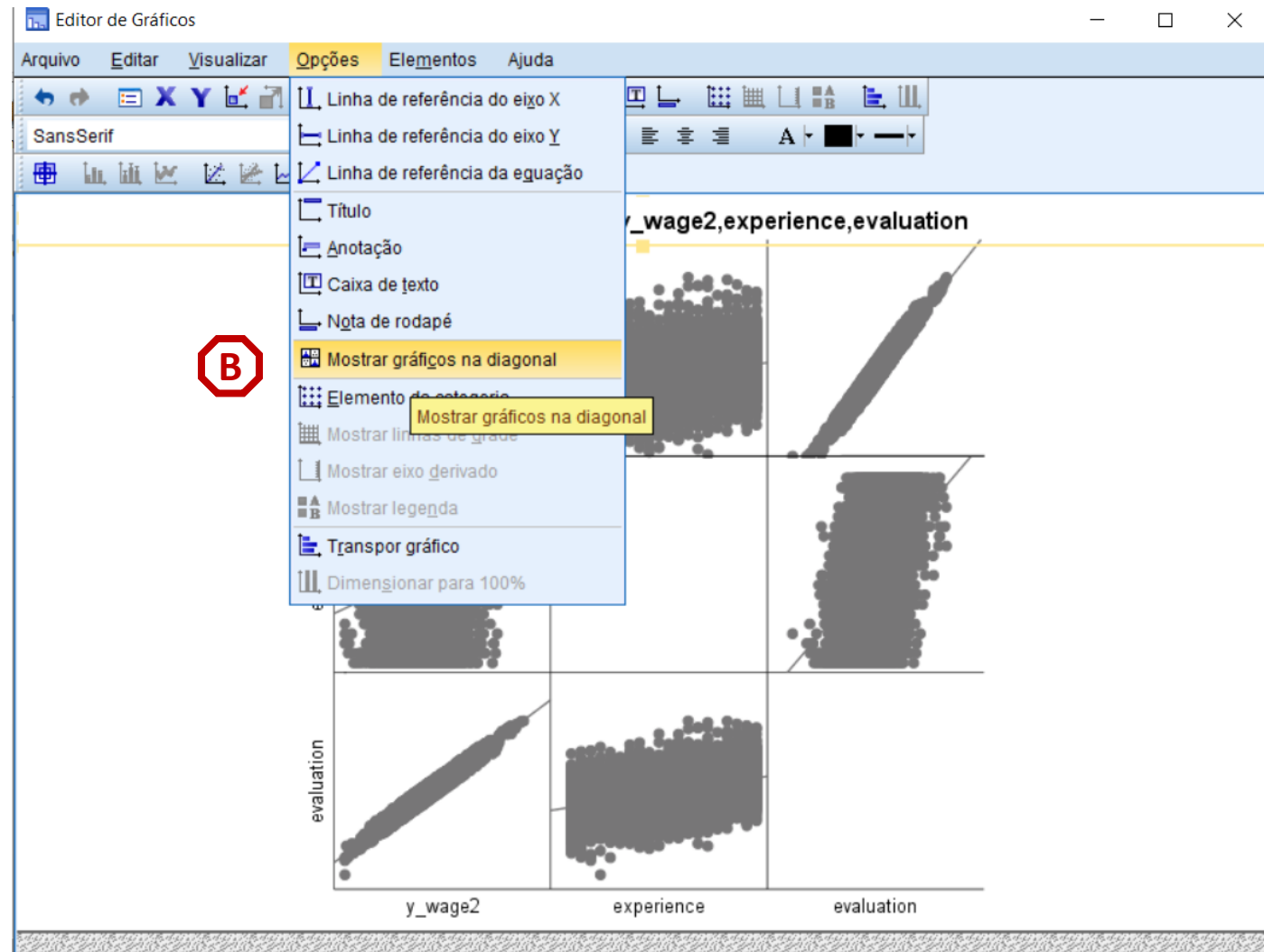
## Matriz de Dispersão

- No Editor de Gráficos...
- Vamos selecionar a opção 'Incluir linha de ajuste no total'
- Vamos querer também selecionar 'Opções' / 'Mostrar gráficos na diagonal'

A

B

B



## Matriz de Dispersão

- No Editor de Gráficos...
- Vamos selecionar a opção 'Incluir linha de ajuste no total'
- Vamos querer também selecionar 'Opções' / 'Mostrar gráficos na diagonal'
- Basta clicar 'Fechar' (e fechar o 'Editor de Gráficos) para vermos o resultado final

A

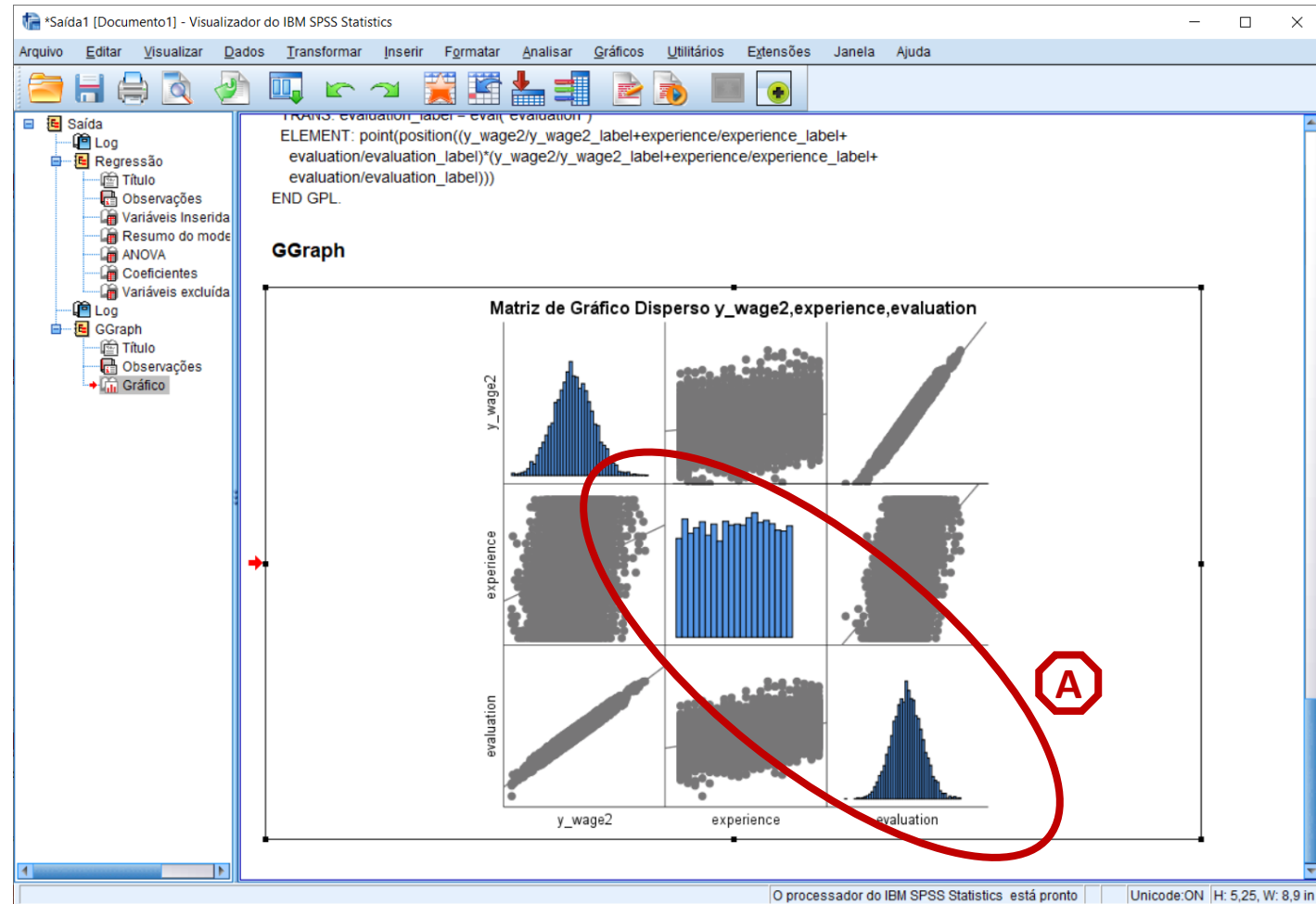
B

C

The screenshot shows the 'Editor de Gráficos' window with a menu bar (Arquivo, Editar, Visualizar, Opções, Elementos, Ajuda) and a toolbar. The main area displays a 'Matriz de Gráfico Dispersão' with three plots: a histogram for 'y\_wage2', a scatter plot for 'experience', and a scatter plot for 'evaluation'. A 'Propriedades' dialog box is open, showing options for 'Tamanho do gráfico', 'Preenchimento e borda', 'Gráficos na diagonal', and 'Variáveis'. The 'Preenchimento e borda' tab is active, showing a color palette and a 'Fechar' button circled in red. The 'Estilo da borda' section has 'Ponderação' set to 1 and 'Estilo' set to 'Agrupado'. The 'Fechar' button is also circled in red.

### Matriz de Dispersão

- Esta Matriz de dispersão permite-nos ver:
- A distribuição das IVs



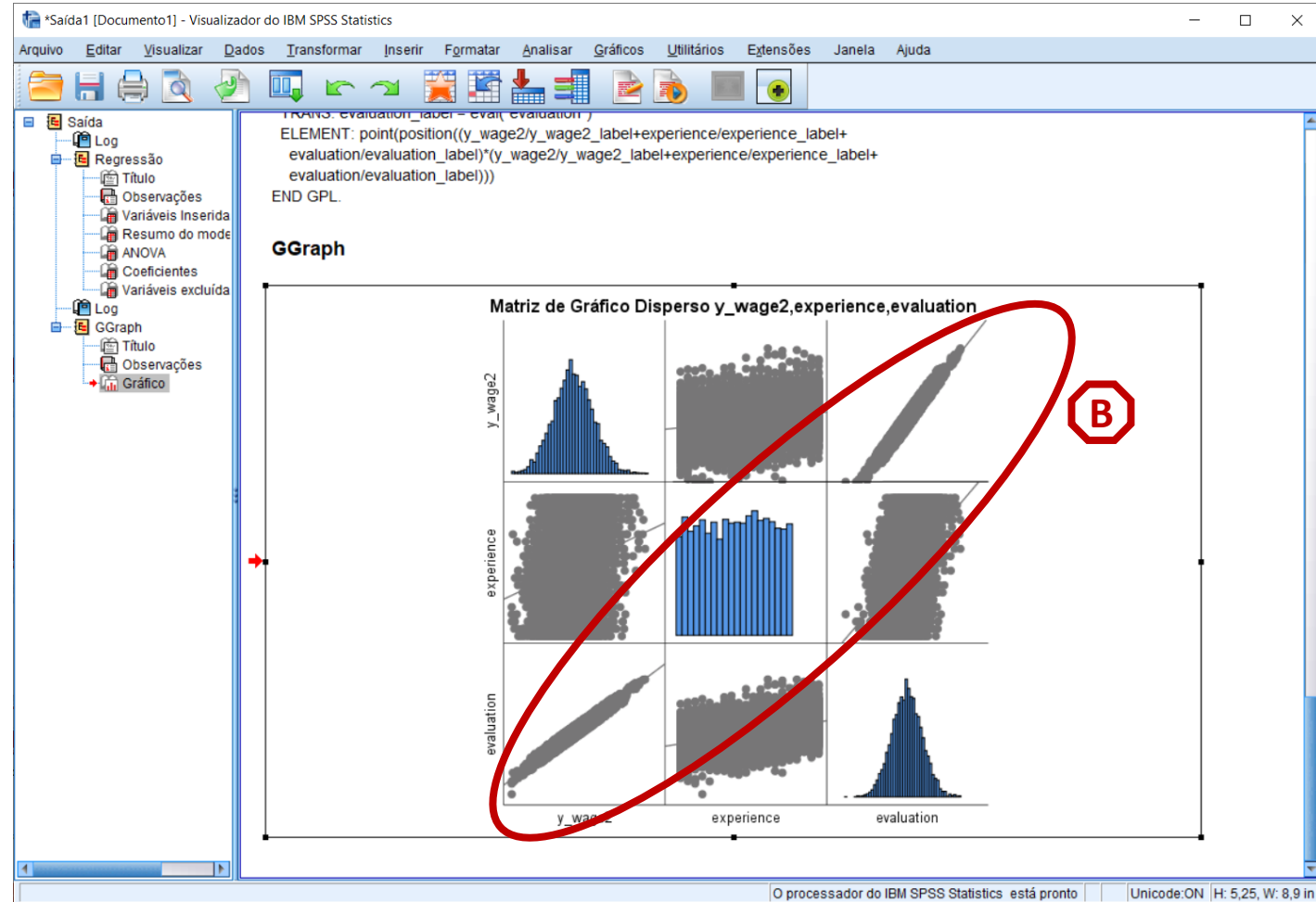
## Matriz de Dispersão

- Esta Matriz de dispersão permite-nos ver:
- A distribuição das IVs
- E a forma como se correlacionam com a VD:
  - Há uma correlação forte entre 'y\_wage2' e a variável 'evaluation'

A

B

C



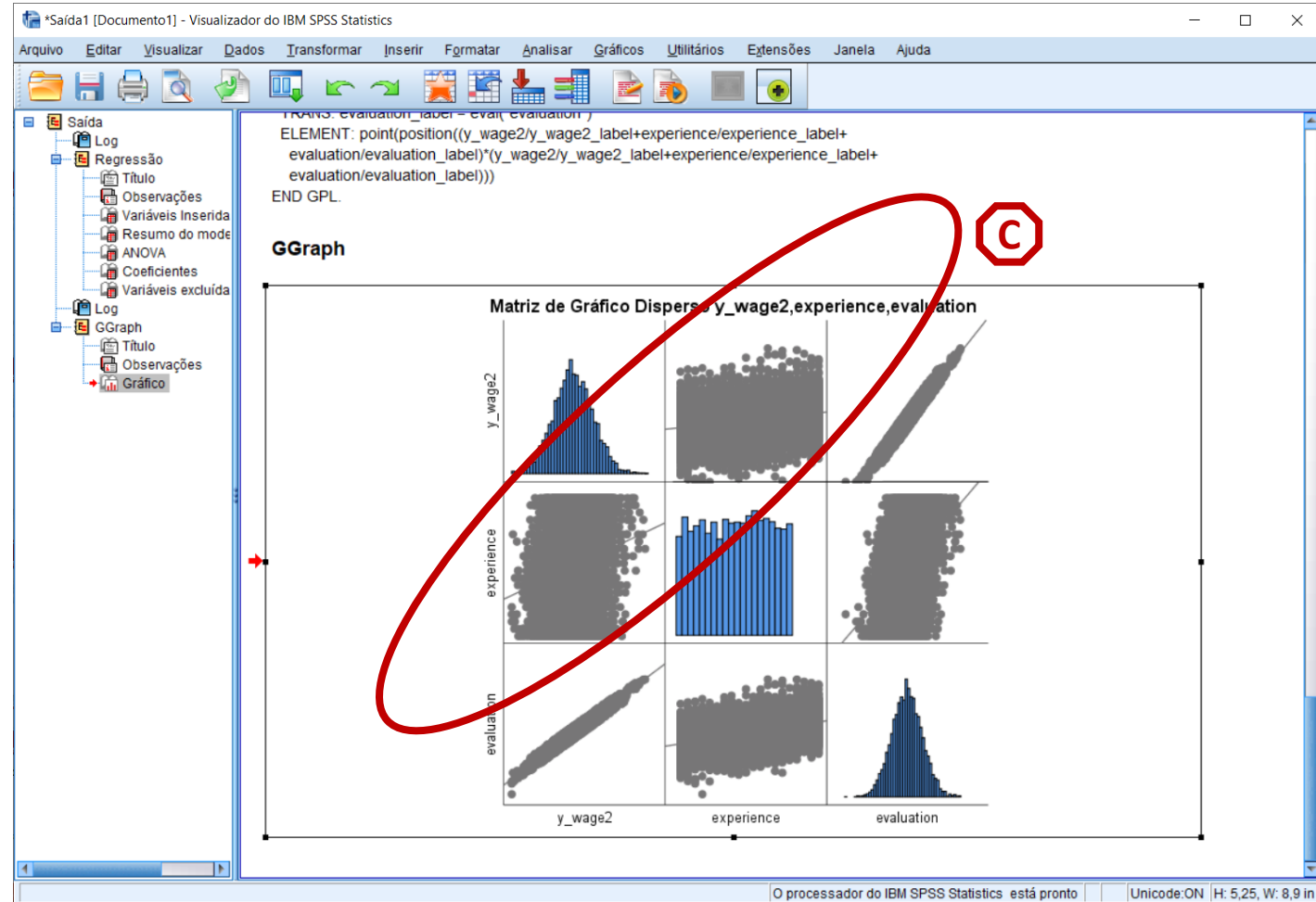
## Matriz de Dispersão

- Esta Matriz de dispersão permite-nos ver:
- A distribuição das IVs
- E a forma como se correlacionam com a VD:
  - Há uma correlação positiva, forte entre 'y\_wage2' e a variável 'evaluation'
  - Uma correlação positiva, mas não tão forte, entre 'y\_wage2' e a variável 'experience'

A

B

C





# Implementação do Modelo de Regressão Linear

*4. Definir um modelo de regressão linear (IV contínuas)*

## Regressão Linear

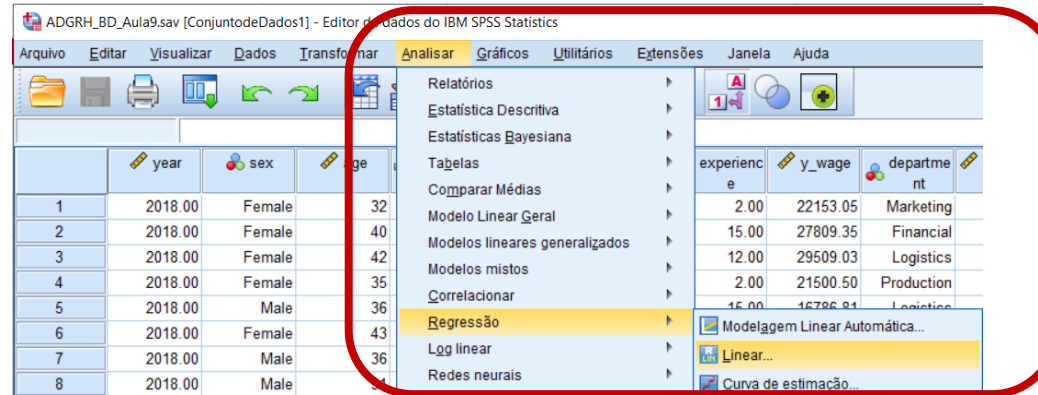
- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Dependente'

Exercício: Colocar as variáveis 'experience' e 'evaluation' na caixa 'Independente(s)'

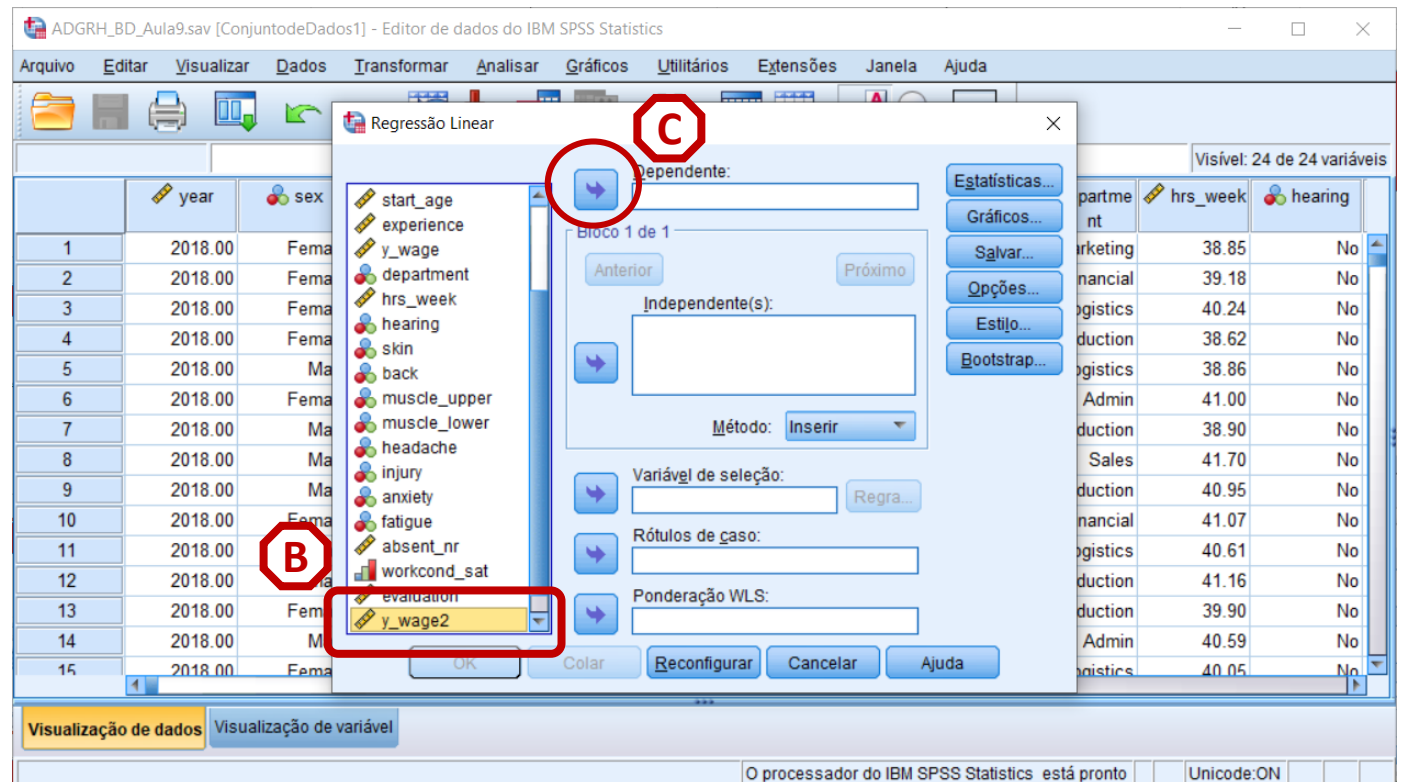
A

B

C



A



B

C

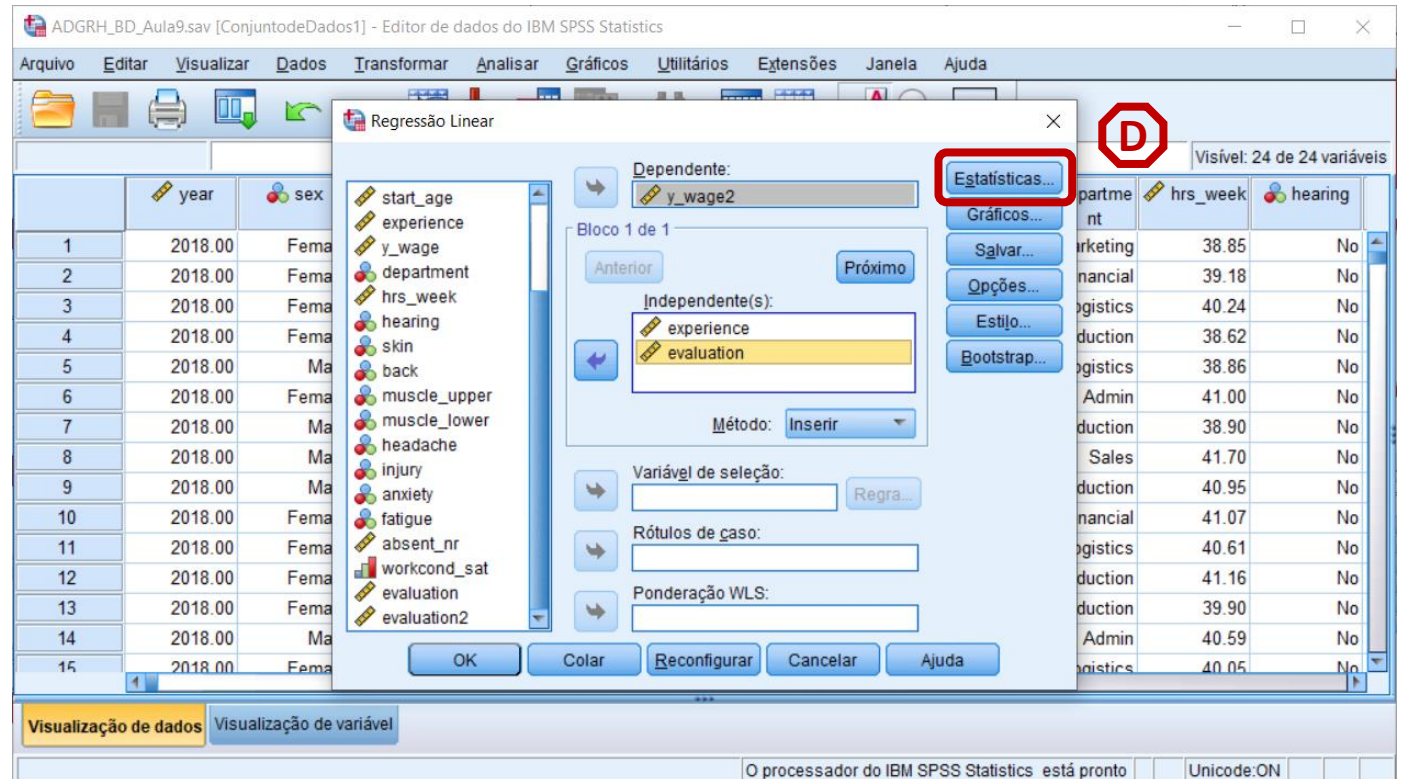
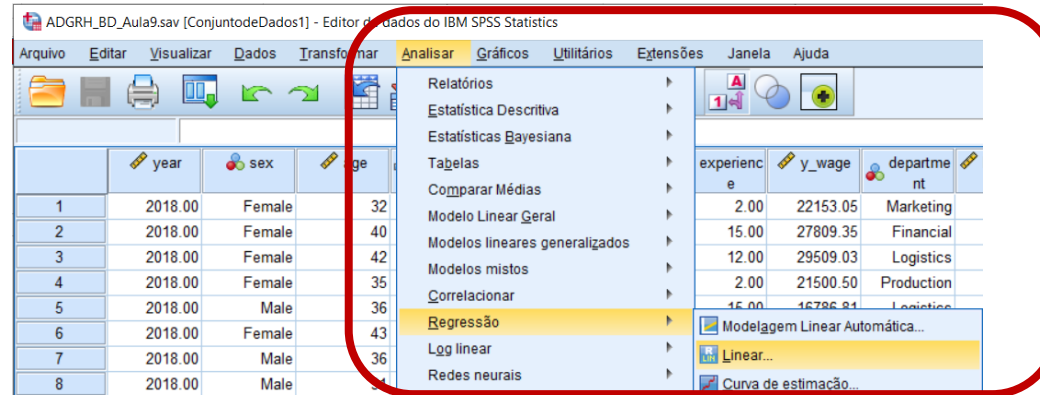


### Regressão Linear

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Dependente'

Exercício: Colocar as variáveis 'experience' e 'evaluation' na caixa 'Independente(s)'

- Selecionar botão 'Estatísticas'



## Regressão Linear

- Seleccionar 'Analisar' / 'Regressão' / 'Linear'
- Seleccionar a variável 'y\_wage2'

- Colocar na caixa 'Dependente'

**Exercício:** Colocar as variáveis 'experience' e 'evaluation' na caixa 'Independente(s)'

- Seleccionar botão 'Estatísticas'
- Seleccionar 'Estimativas'
- Seleccionar 'Ajuste do modelo'
- Seleccionar 'Continuar'/'OK'

A

B

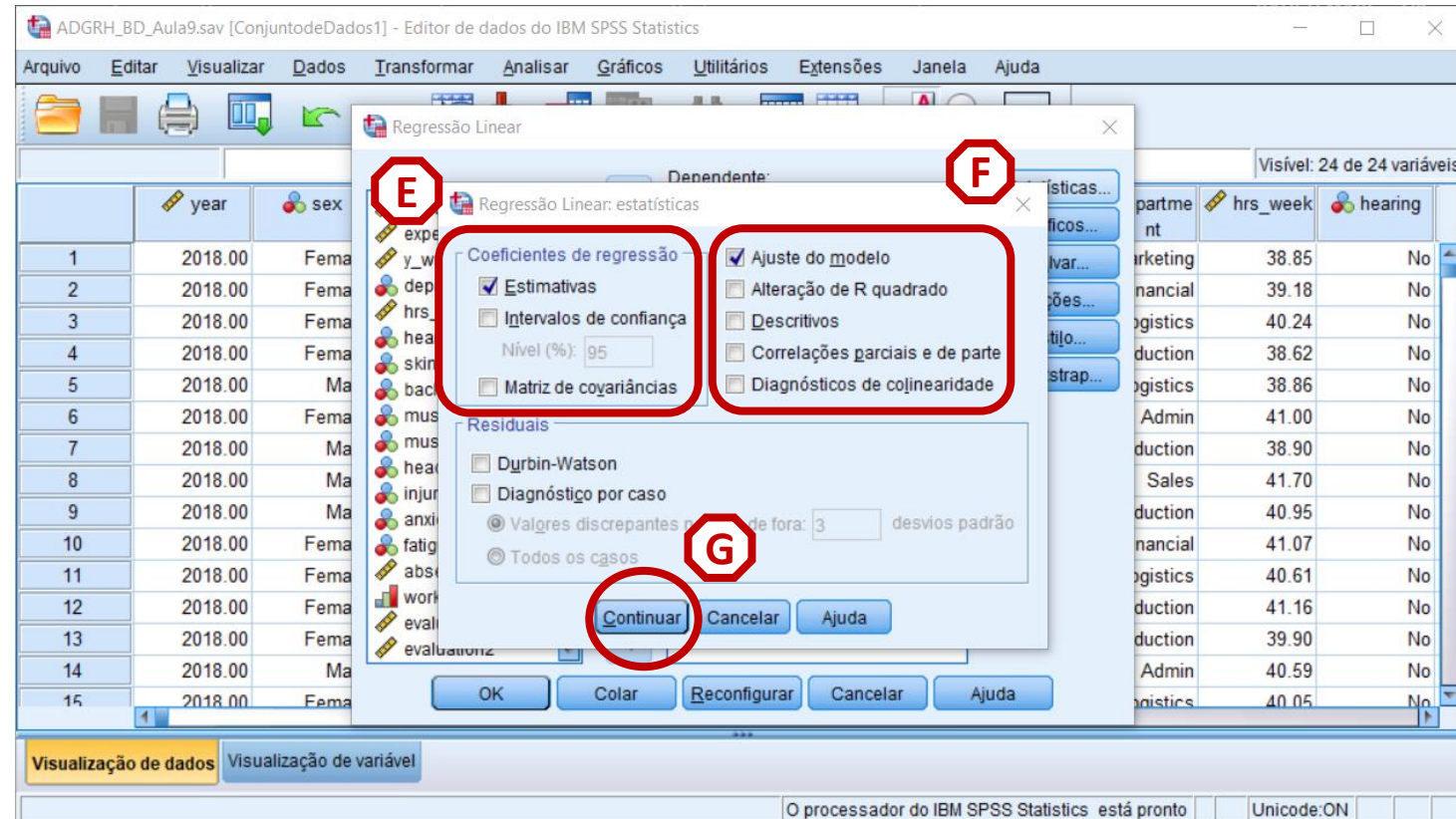
C

D

E

F

G



ADGRH\_BD\_Aula9.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Regressão Linear

Dependente: y\_wage2

Independente(s): experience, evaluation

Regressão Linear: estatísticas

Coeficientes de regressão:

- Estimativas
- Intervalos de confiança
- Matriz de covariâncias

Nível (%): 95

Ajuste do modelo:

- Ajuste do modelo
- Alteração de R quadrado
- Descritivos
- Correlações parciais e de parte
- Diagnósticos de colinearidade

Residuais:

- Durbin-Watson
- Diagnóstico por caso
- Valgres discrepantes: de fora: 3 desvios padrão
- Todos os casos

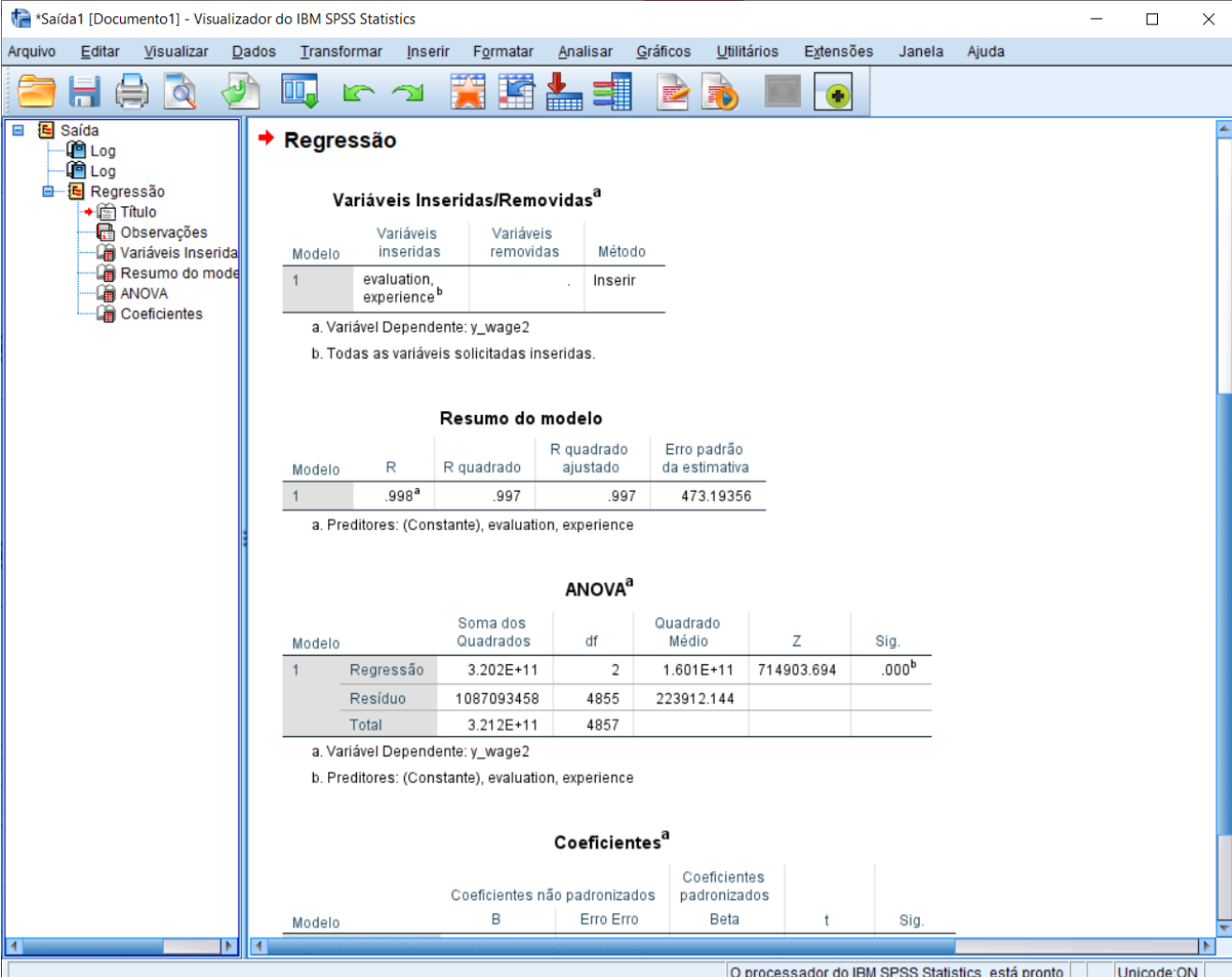
Continuar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Regressão Linear

- O resultado é publicado no 'Visualizador de Resultados'



\*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

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

➔ **Regressão**

**Variáveis Inseridas/Removidas<sup>a</sup>**

Modelo	Variáveis inseridas	Variáveis removidas	Método
1	evaluation, experience <sup>b</sup>	.	Inserir

a. Variável Dependente: y\_wage2  
b. Todas as variáveis solicitadas inseridas.

**Resumo do modelo**

Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa
1	.998 <sup>a</sup>	.997	.997	473.19356

a. Preditores: (Constante), evaluation, experience

**ANOVA<sup>a</sup>**

Modelo		Soma dos Quadrados	df	Quadrado Médio	Z	Sig.
1	Regressão	3.202E+11	2	1.601E+11	714903.694	.000 <sup>b</sup>
	Resíduo	1087093458	4855	223912.144		
	Total	3.212E+11	4857			

a. Variável Dependente: y\_wage2  
b. Preditores: (Constante), evaluation, experience

**Coefficientes<sup>a</sup>**

Modelo	Coefficientes não padronizados		Coefficientes padronizados		t	Sig.
	B	Erro Erro	Beta			

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

## Regressão Linear

O que é que esta tabela nos diz?

- Qual é a variável dependente?

*'y\_wage2'*

- Quais são as variáveis independentes?

*experiência*

*avaliação dos trabalhadores*

- Alguma variável foi excluída?

*Não*

Variáveis Inseridas/Removidas<sup>a</sup>

Modelo	Variáveis inseridas	Variáveis removidas	Método
1	evaluation, experience <sup>b</sup>	.	Inserir

a. Variável Dependente: y\_wage2

b. Todas as variáveis solicitadas inseridas.

## Regressão Linear

- O nosso modelo é estatisticamente significativo?

O valor Z é maior que 0

→ pelo menos uma das IVs tem uma relação estatisticamente significativa com a DV

O valor Sig. é menor que 0.01

→ O modelo é estatisticamente significativa com um grau de confiança a 99%

		ANOVA <sup>a</sup>			<b>A</b>	<b>B</b>
Modelo		Soma dos Quadrados	df	Quadrado Médio	Z	Sig.
1	Regressão	3.202E+11	2	1.601E+11	714903.694	.000 <sup>b</sup>
	Resíduo	1087093458	4855	223912.144		
	Total	3.212E+11	4857			

a. Variável Dependente: y\_wage2

b. Preditores: (Constante), evaluation, experience

**A**

**B**

## Regressão Linear

- Qual é o poder explicativo do nosso?

O  $R^2$  é de .997

→ *O modelo explica 97% da variação dos salários na organização*

**!! Este tipo de resultado reflecte o facto de esta ser uma base de dados sintética !!**

A

**Resumo do modelo**

Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa
1	.998 <sup>a</sup>	.997	.997	473.19356

a. Preditores: (Constante), evaluation, experience

## Regressão Linear

- O que diz a tabela sobre o efeito das variáveis independentes?

O valor Sig. é menor que 0.01

→ As variáveis 'experiência' e 'avaliação' têm uma relação estatisticamente significativa com a DV, com um grau de confiança a 99%

O efeito da variável 'evaluation' sobre o salário anual dos trabalhadores é superior ao efeito da variável 'experience'.

A

B

**Coeficientes<sup>a</sup>**

Modelo		Coeficientes não padronizados		Coeficientes padronizados	t	Sig.
		B	Erro Erro	Beta		
1	(Constante)	-146805.438	145.595		-1008.312	.000
	experience	-341.171	1.316	-.240	-259.307	.000
	evaluation	3354.724	2.879	1.078	1165.235	.000

a. Variável Dependente: y\_wage2

## Regressão Linear

- O que diz a tabela sobre o efeito das variáveis independentes?

Por cada ano adicional de experiência o valor do salário (anual) desce 341 Euros.

Por cada valor adicional na avaliação o valor do salário (anual) aumenta 3355 Euros.

C

D

**Coeficientes<sup>a</sup>**

Modelo		Coeficientes não padronizados		Coeficientes padronizados		Sig.
		B	Erro Erro	Beta	t	
1	(Constante)	-146805.438	145.595		-1008.312	.000
	experience	-341.171	1.316	-.240	-259.307	.000
	evaluation	3354.724	2.879	1.078	1165.235	.000

a. Variável Dependente: y\_wage2



## Regressão Linear

- O que diz a tabela sobre o efeito das variáveis independentes?

Por cada ano adicional de experiência o valor do salário (anual) desce 341 Euros.

Por cada valor adicional na avaliação o valor do salário (anual) aumenta 3355 Euros.

C

D

**Coefficientes<sup>a</sup>**

Modelo		Coeficientes não padronizados		Coeficientes padronizados	t	Sig.
		B	Erro Erro	Beta		
1	(Constante)	-146805.438	145.595		-1008.312	.000
	experience	-341.171	1.316	-.240	-259.307	.000
	evaluation	3354.724	2.879	1.078	1165.235	.000

a. Variável Dependente: y\_wage2

O que é que isto mostra?

- Não basta apenas buscar por variáveis estatisticamente significativas!!
- Temos de olhar ao efeito da variável!!
- Neste caso, convém incluir mais variáveis no nosso modelo!!

# Implementação do Modelo de Regressão Linear

## *4. Adicionar IVs nominais*

## Adicionar IVs nominais

- **Objectivo:**
  - Introduzir a variável 'sex' no modelo de regressão que explica a variação dos salários na organização

**Mas antes... precisamos de criar uma dummy variável 'sex2', de modo a que esta assuma os valores 0 (mulheres) e 1 (homens)**

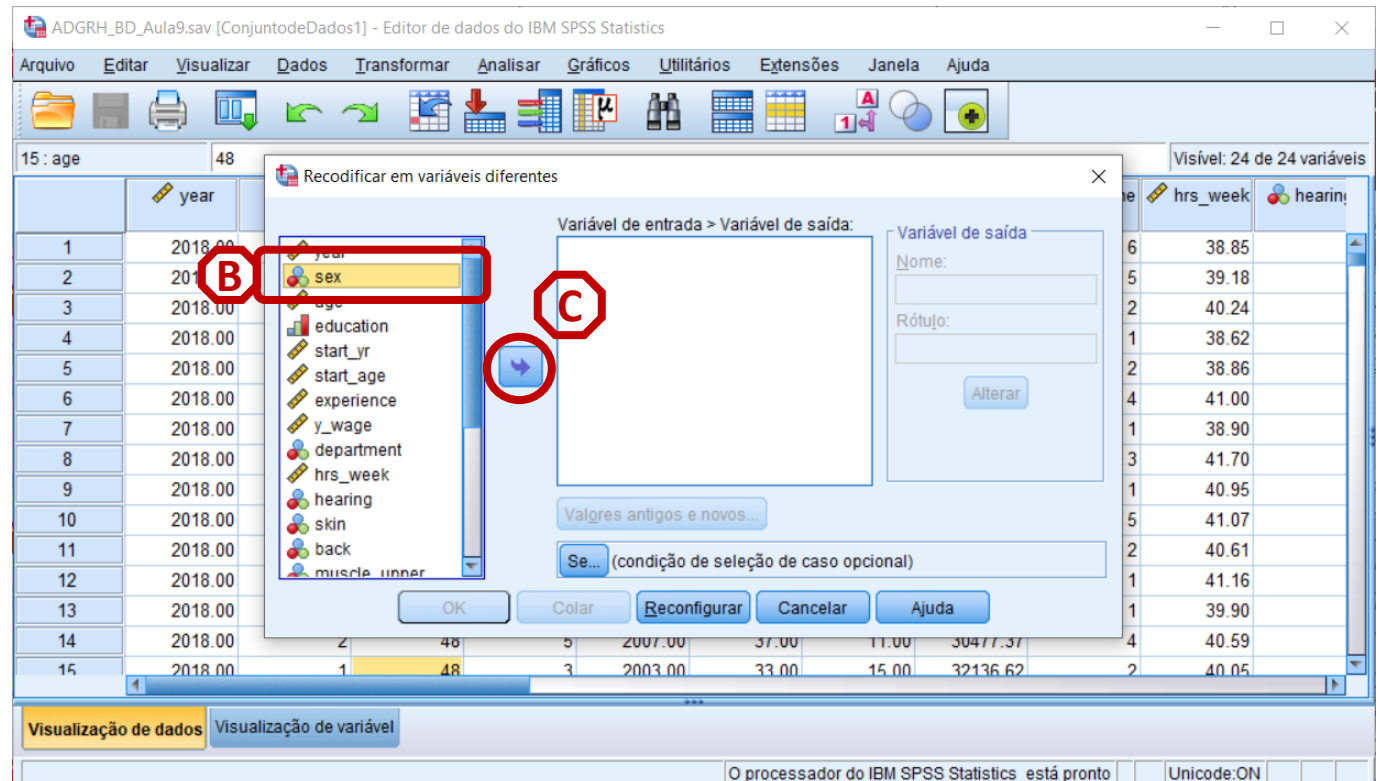
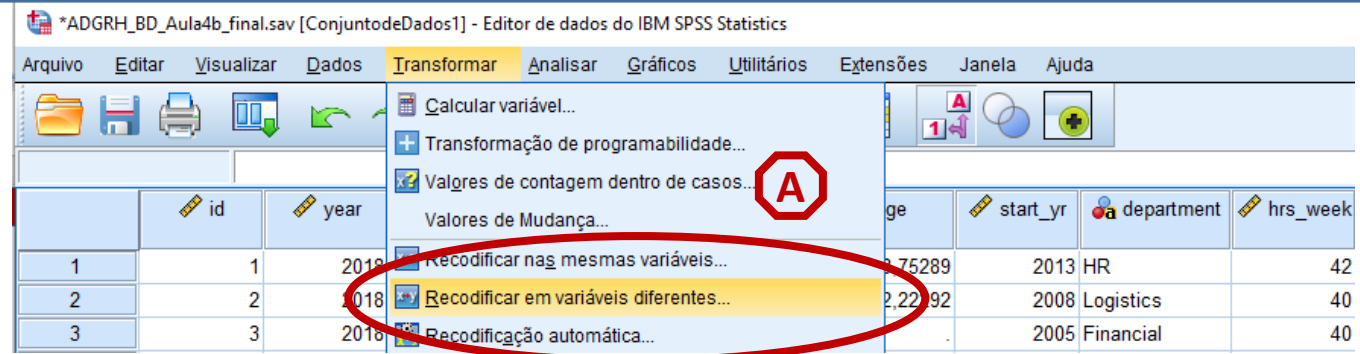
### Adicionar IVs nominais

- Selecionar ‘Transformar’/’Recodificar em variáveis diferentes’
- Selecionar a variável ‘sex’...
- ... e colocar na caixa da ‘variável de entrada’

A

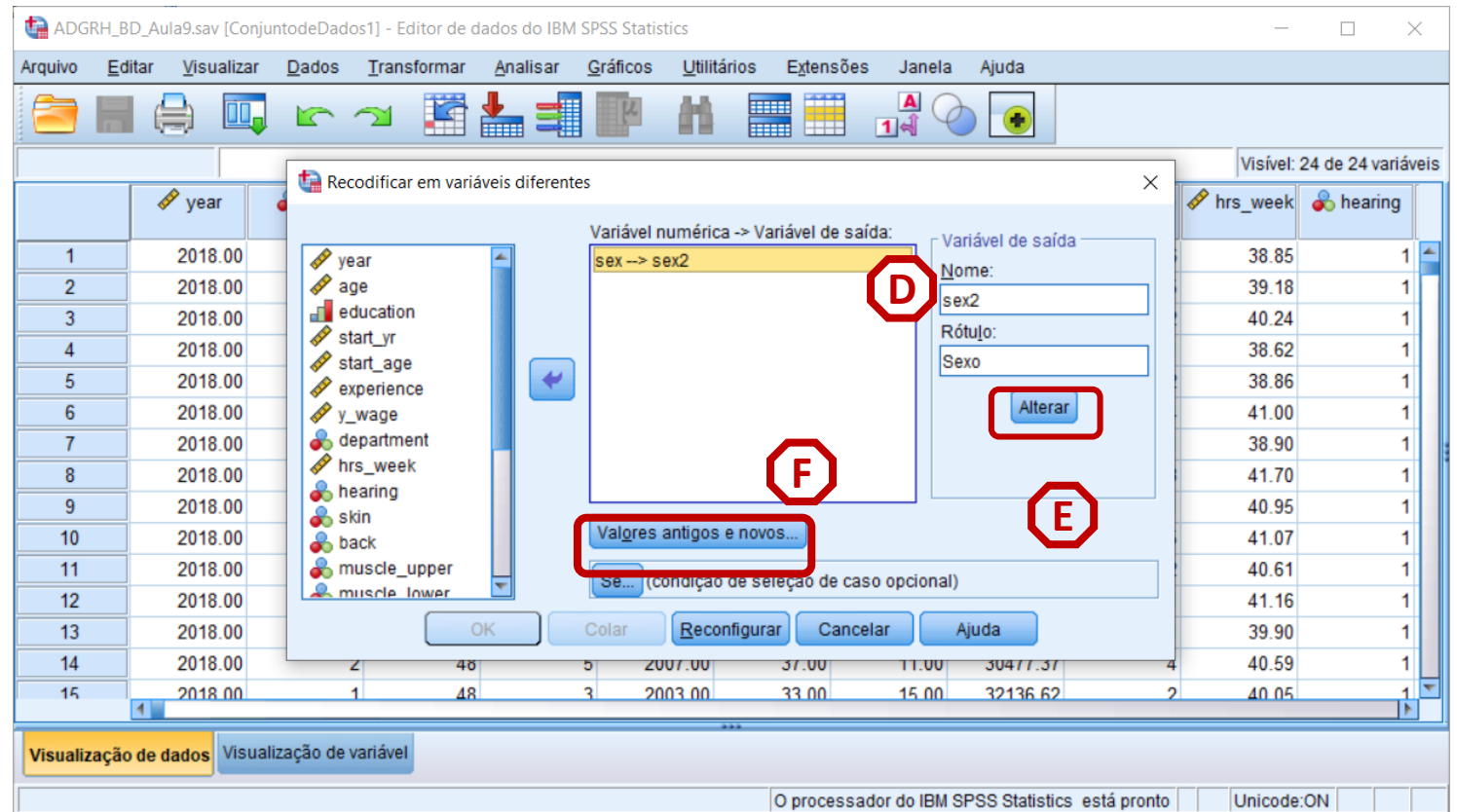
B

C



### Adicionar IVs nominais

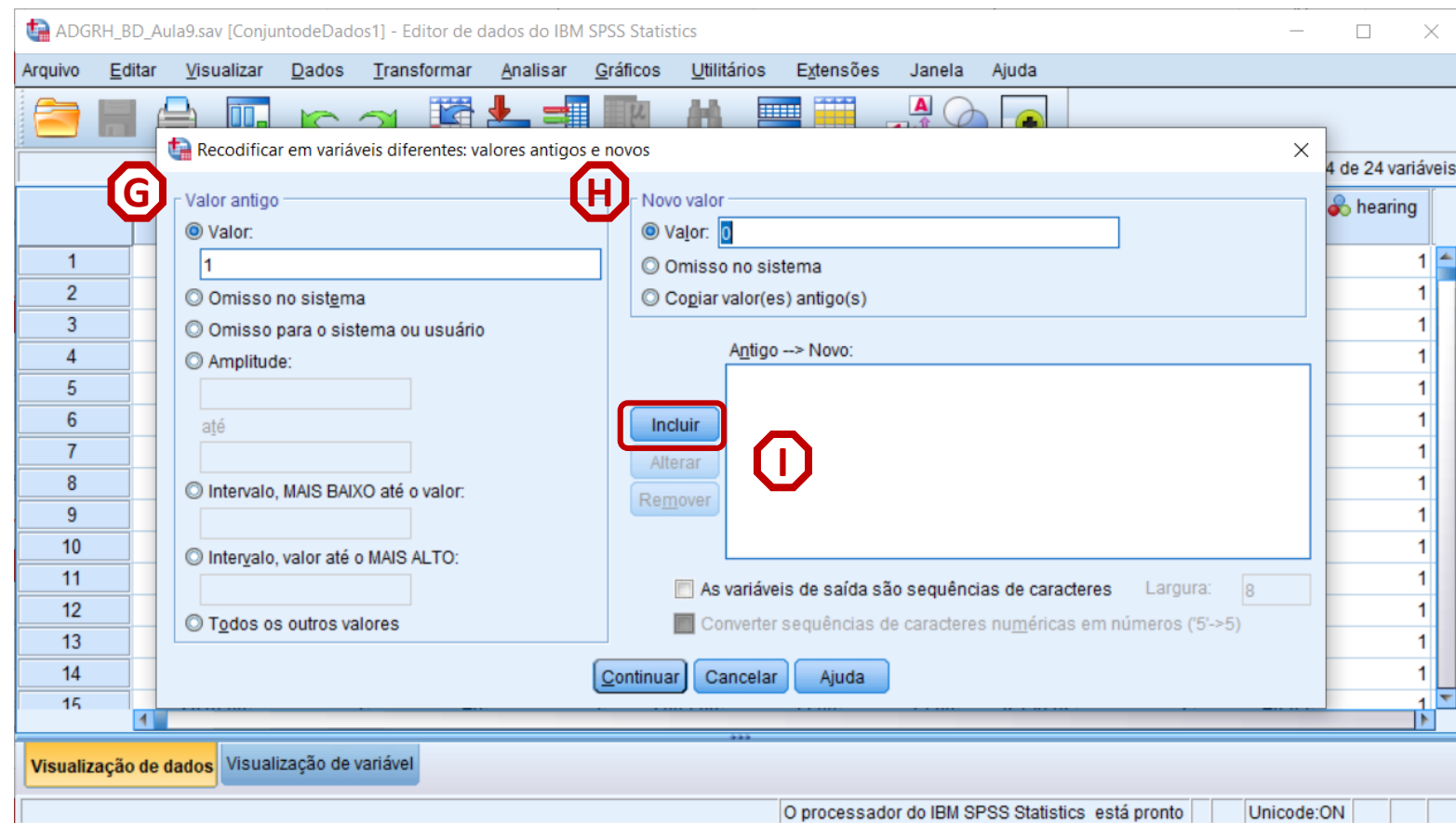
- Seleccionar ‘Transformar’/’Recodificar em variáveis diferentes’ **A**
- Seleccionar a variável ‘sex’... **B**
- ... e colocar na caixa da ‘variável de entrada’ **C**
- Definir o nome da nova variável (‘sex2’) e o rótulo da variável (‘Sexo’) **D**
- Seleccionar o botão ‘Alterar’ **E**
- Seleccionar ‘Valores antigo e novo’ **F**



## Adicionar IVs nominais

- Vamos definir que ao valor '1' na variável 'sex' **G**
- Corresponde o valor '0', na nova variável ('sex2') **H**
- Selecionar 'Incluir' **I**

**Exercício:** Associar valor '1' na variável 'sex2' ao valor '2' na variável original ('sex')

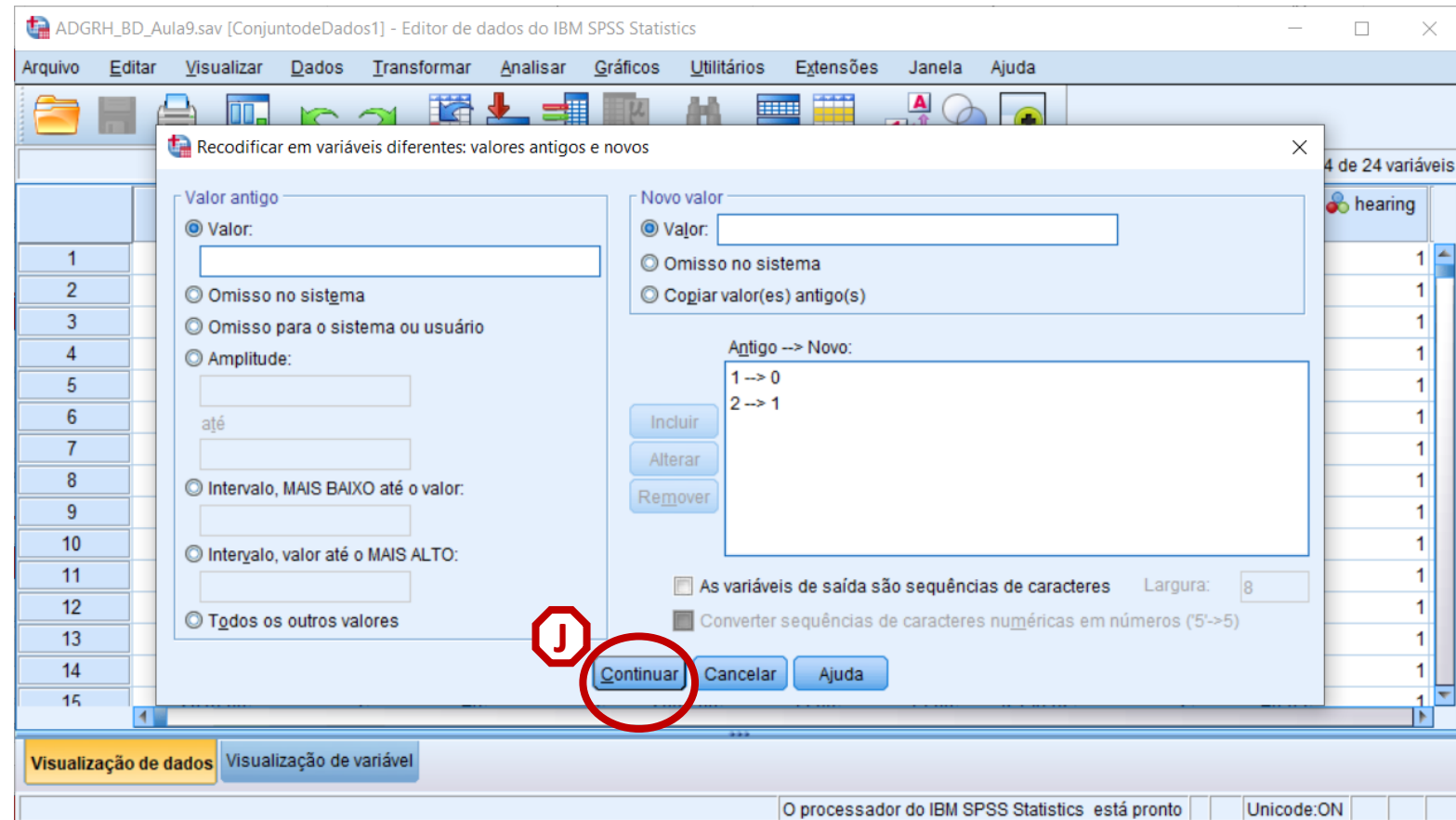


## Adicionar IVs nominais

- Vamos definir que ao valor '1' na variável 'sex' **G**
- Corresponde o valor '0', na nova variável ('sex2') **H**
- Selecionar 'Incluir' **I**

**Exercício:** Associar valor '1' na variável 'sex2' ao valor '2' na variável original ('sex')

- Selecionar 'Continuar' / 'OK' **J**



### Adicionar IVs nominais

- Vamos verificar se a recodificação foi bem sucedida

\*ADGRH\_BD\_Aula9.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : sex2 .0 Visível: 25 de 25 variáveis

	fatigue	absent_nr	workcond_sat	evaluation	y_wage2	evaluation_2	sex2	var	var	var	var
1	1	9	3	49.08	17529.60	55.17	.00				
2	1	10	2	52.58	24843.10	51.75	.00				
3	1	15	2	53.80	29874.72	54.79	.00				
4	1	10	2	50.42	20994.46	58.30	.00				
5	1	11	4	48.43	10044.01	49.89	1.00				
6	1	10	2	54.34	32321.53	52.74	.00				
7	1	6	3	49.65	18757.44	47.60	1.00				
8	1	10	4	46.84	9106.36	52.65	1.00				
9	1	8	2	50.76	20432.69	48.48	1.00				
10	1	17	3	53.33	29327.90	52.40	.00				
11	1	14	2	50.59	18223.72	47.21	.00				
12	1	7	2	54.64	29668.47	48.39	.00				
13	1	13	3	56.64	35658.63	50.19	.00				
14	1	7	3	55.36	35733.01	40.86	1.00				
15	1	11	3	56.09	36160.52	45.36	.00				

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON



# ANÁLISE DE DADOS EM GRH

## Aula 9: O modelo de regressão linear

### Adicionar IVs nominiais

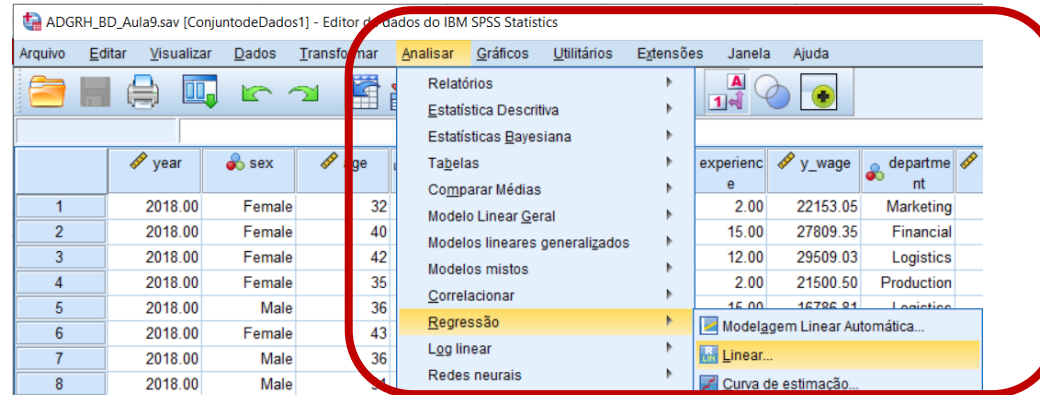
- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Dependente'

**Exercício:** Colocar as variáveis 'experience' e 'evaluation' na caixa 'Independente(s)'

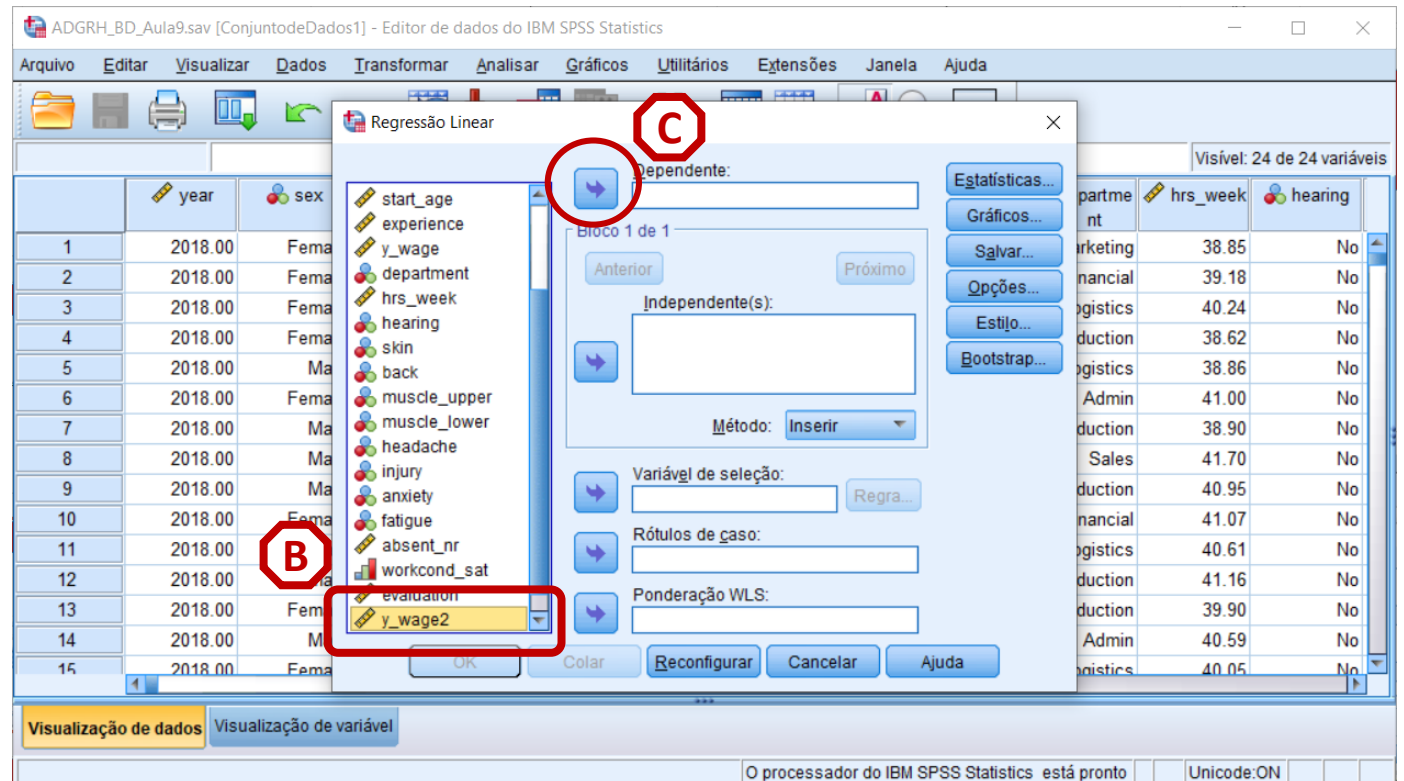
A

B

C



A



C

B

### Adicionar IVs nominais

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
  - Selecionar a variável 'y\_wage2'
  - Colocar na caixa 'Dependente'
  - Colocar na caixa 'Independente(s)'
  - Colocar na caixa 'Independente(s)'
- Exercício: Colocar as variáveis 'experience' e 'evaluation' na caixa 'Independente(s)'
- Selecionar a variável 'sex2'
  - Colocar na caixa 'Independente(s)'

A

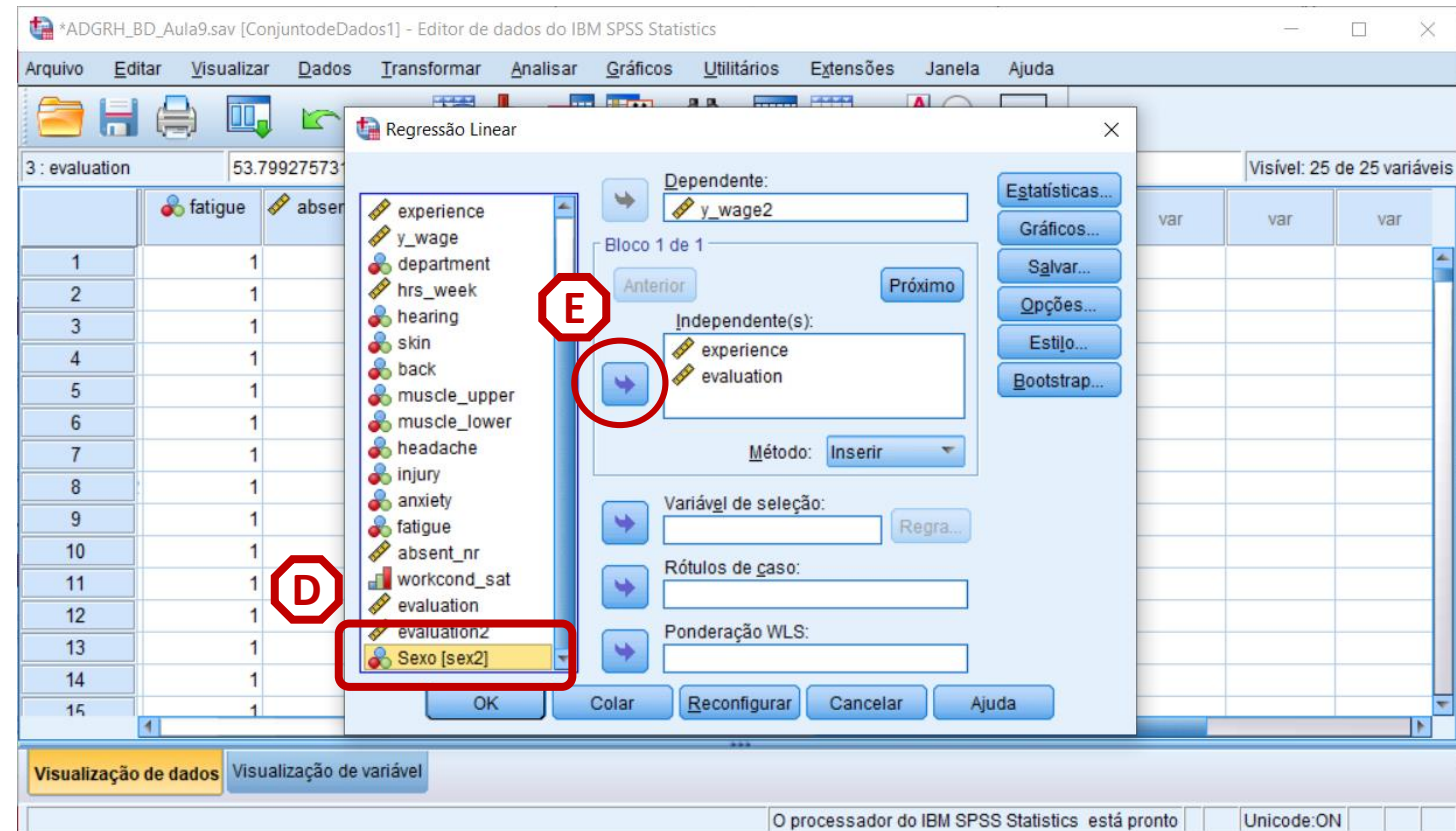
B

C

D

D

E



\*ADGRH\_BD\_Aula9.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Regressão Linear

Dependente: y\_wage2

Bloco 1 de 1

Independente(s): experience, evaluation

Método: Inserir

Variável de seleção:

Rótulos de caso:

Ponderação WLS:

OK Colar Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

### Adicionar IVs nominais

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Dependente'
- Colocar na caixa 'Independente(s)'
- Colocar na caixa 'Independente(s)'
- Selecionar a variável 'sex2'
- Colocar na caixa 'Independente(s)'
- Selecionar botão 'Estatísticas'

A

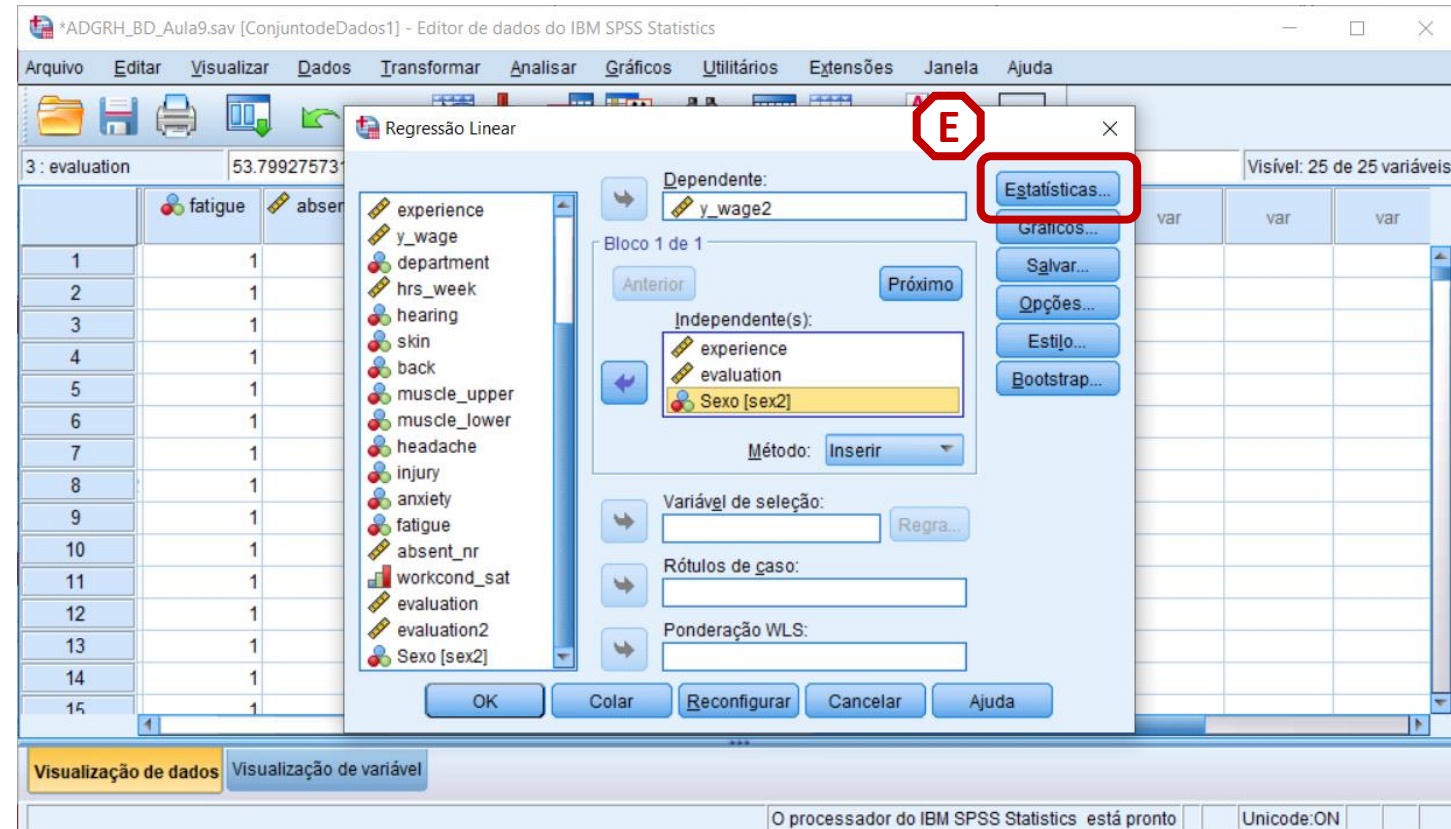
B

C

D

E

F



## Adicionar IVs nominais

- Selecionar 'Estimativas' **G**
- Selecionar 'Ajuste do modelo' **H**
- Selecionar 'Continuar'/'OK' **I**

ADGRH\_BD\_Aula9.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Regressão Linear

Dependente: **H**

Regressão Linear: estatísticas **G**

Estimativas  
 Intervalos de confiança  
Nível (%): 95  
 Matriz de covariâncias

Ajuste do modelo  
 Alteração de R quadrado  
 Descritivos  
 Correlações parciais e de parte  
 Diagnósticos de colinearidade

Residuais  
 Durbin-Watson  
 Diagnóstico por caso  
 Valgres discrepantes r de fora: 3 desvios padrão  
 Todos os casos **I**

**Continuar** Cancelar Ajuda

OK Colar Reconfigurar Cancelar Ajuda

	year	sex	hrs_week	hearing
1	2018.00	Fema		
2	2018.00	Fema		
3	2018.00	Fema		
4	2018.00	Fema		
5	2018.00	Ma		
6	2018.00	Fema		
7	2018.00	Ma		
8	2018.00	Ma		
9	2018.00	Ma		
10	2018.00	Fema		
11	2018.00	Fema		
12	2018.00	Fema		
13	2018.00	Fema		
14	2018.00	Ma		
15	2018.00	Fema		

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

### Adicionar IVs nominais

- O que diz a tabela sobre o efeito das variáveis independentes?

O valor Sig. é menor que 0.05

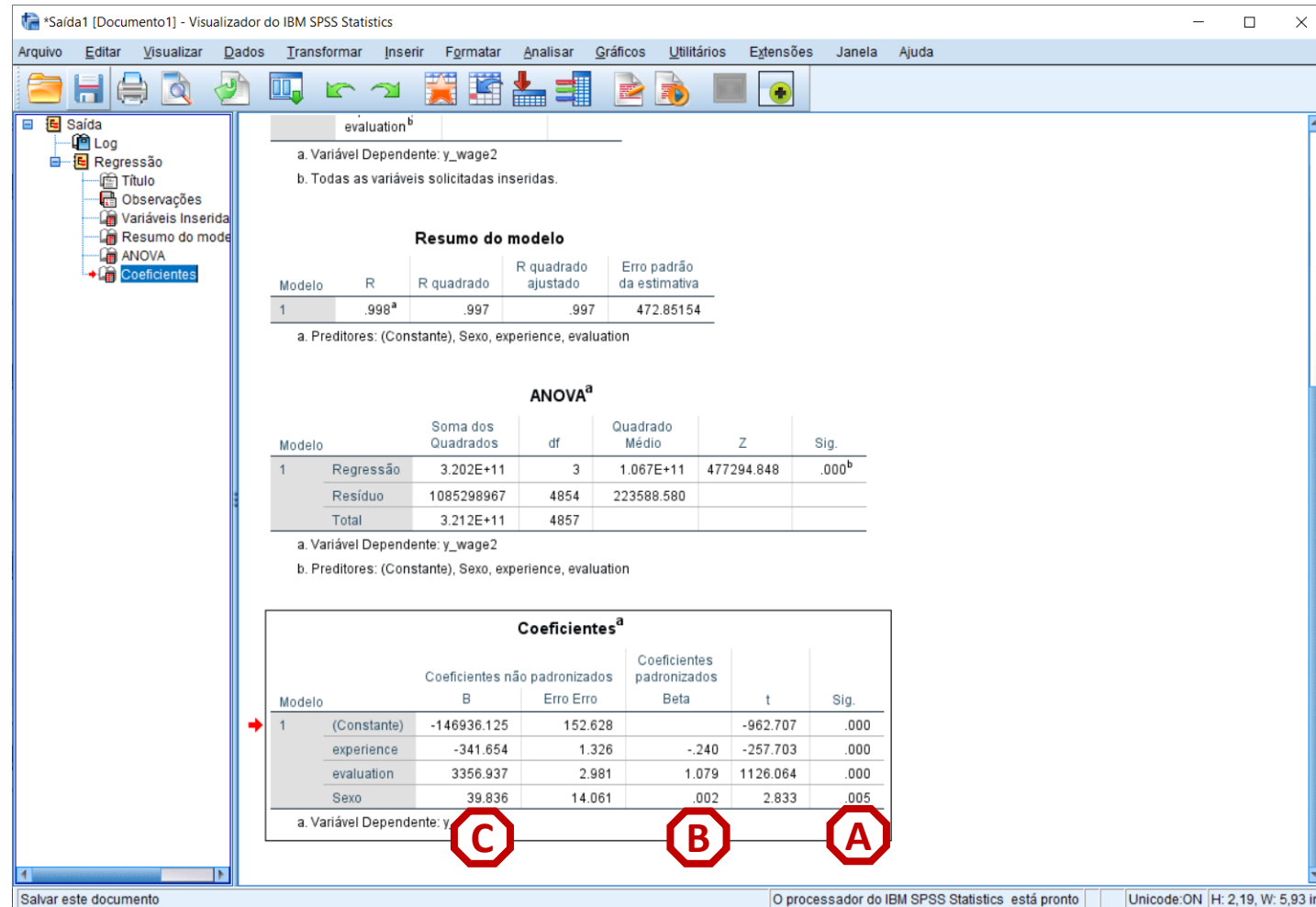


→ A variáveis 'sex' tem uma relação estatisticamente significativa com a DV, com um grau de confiança a 95%

O efeito de ser homem é o menor de todas as variáveis



Ser homem aumenta o salário (anual) em 40 Euros, em relação às mulheres.

Visualizador do IBM SPSS Statistics

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

évaluation<sup>b</sup>

a. Variável Dependente: y\_wage2  
b. Todas as variáveis solicitadas inseridas.

**Resumo do modelo**

Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa
1	.998 <sup>a</sup>	.997	.997	472.85154

a. Preditores: (Constante), Sexo, experience, evaluation

**ANOVA<sup>a</sup>**

Modelo		Soma dos Quadrados	df	Quadrado Médio	Z	Sig.
1	Regressão	3.202E+11	3	1.067E+11	477294.848	.000 <sup>b</sup>
	Resíduo	1085298967	4854	223588.580		
	Total	3.212E+11	4857			

a. Variável Dependente: y\_wage2  
b. Preditores: (Constante), Sexo, experience, evaluation

**Coefficientes<sup>a</sup>**

Modelo		Coefficients não padronizados		Coefficients padronizados		Sig.
		B	Erro Erro	Beta	t	
1	(Constante)	-146936.125	152.628		-962.707	.000
	experience	-341.654	1.326	-.240	-257.703	.000
	evaluation	3356.937	2.981	1.079	1126.064	.000
	Sexo	39.836	14.061	.002	2.833	.005

a. Variável Dependente: y

Annotations: A (Sig. < 0.05), B (Sexo), C (Sexo)

Salvar este documento | O processador do IBM SPSS Statistics está pronto | Unicode:ON | H: 2,19, W: 5,93 in

## Adicionar IVs nominais

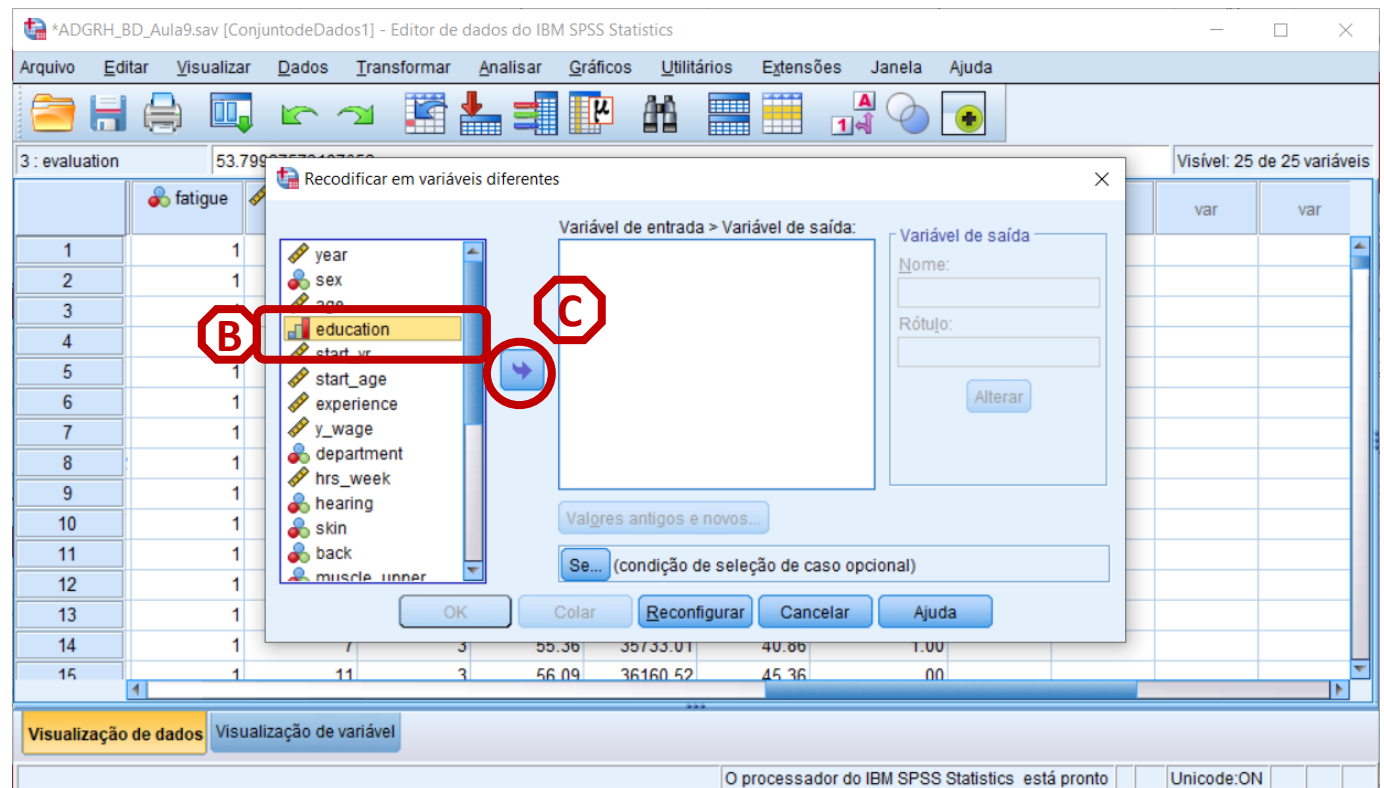
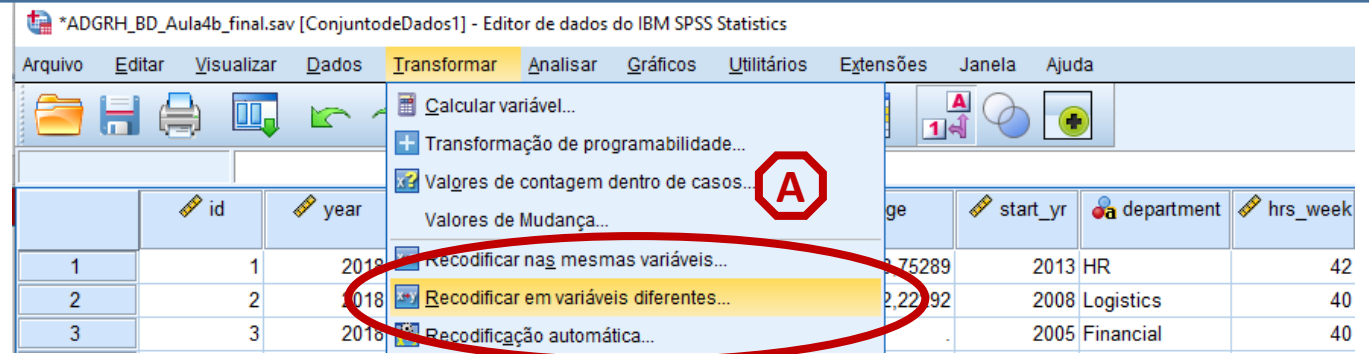
- **Objectivo:**
  - **Introduzir a variável 'education' no modelo de regressão que explica a variação dos salários na organização**

**Infelizmente, o modelo de regressão linear não admite variáveis nominais com mais do que duas categorias (ou variáveis ordinais)...**

**Assim, precisamos de criar uma série de variáveis dummy para cada uma das categorias.**

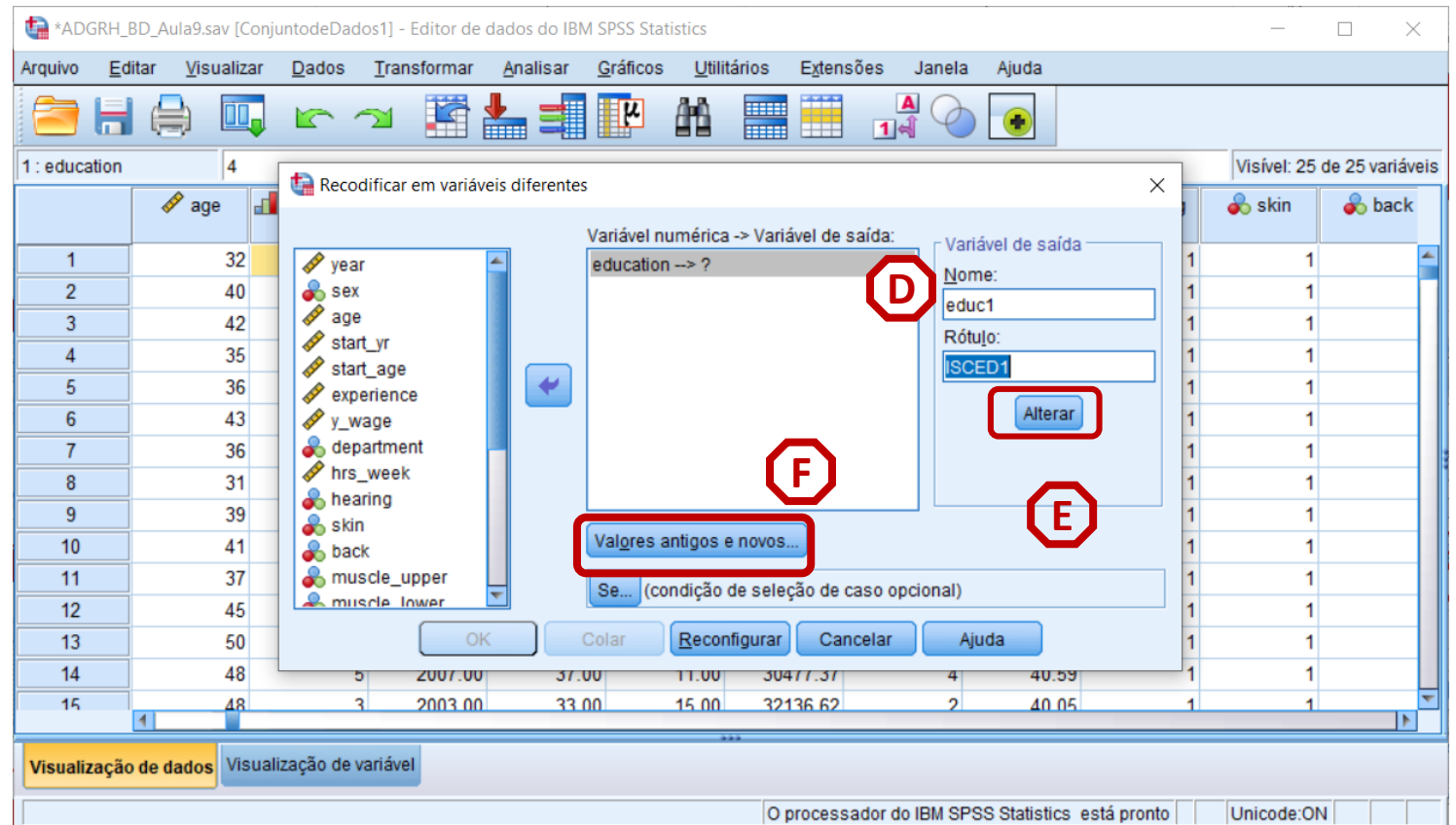
### Adicionar IVs nominais

- Selecionar ‘Transformar’/’Recodificar em variáveis diferentes’ **A**
- Selecionar a variável ‘education’... **B**
- ... e colocar na caixa da ‘variável de entrada’ **C**



### Adicionar IVs nominais

- Seleccionar 'Transformar'/'Recodificar em variáveis diferentes' **A**
- Seleccionar a variável 'education'... **B**
- ... e colocar na caixa da 'variável de entrada' **C**
- Definir o nome da nova variável ('educ1') e o rótulo da variável ('ISCED1') **D**
- Seleccionar o botão 'Alterar' **E**
- Seleccionar 'Valores antigo e novo' **F**



\*ADGRH\_BD\_Aula9.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: education 4

Recodificar em variáveis diferentes

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

Variável de saída

Nome: educ1

Rótulo: ISCED1

Alterar

Valores antigos e novos...

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

OK Colar Reconfigurar Cancelar Ajuda

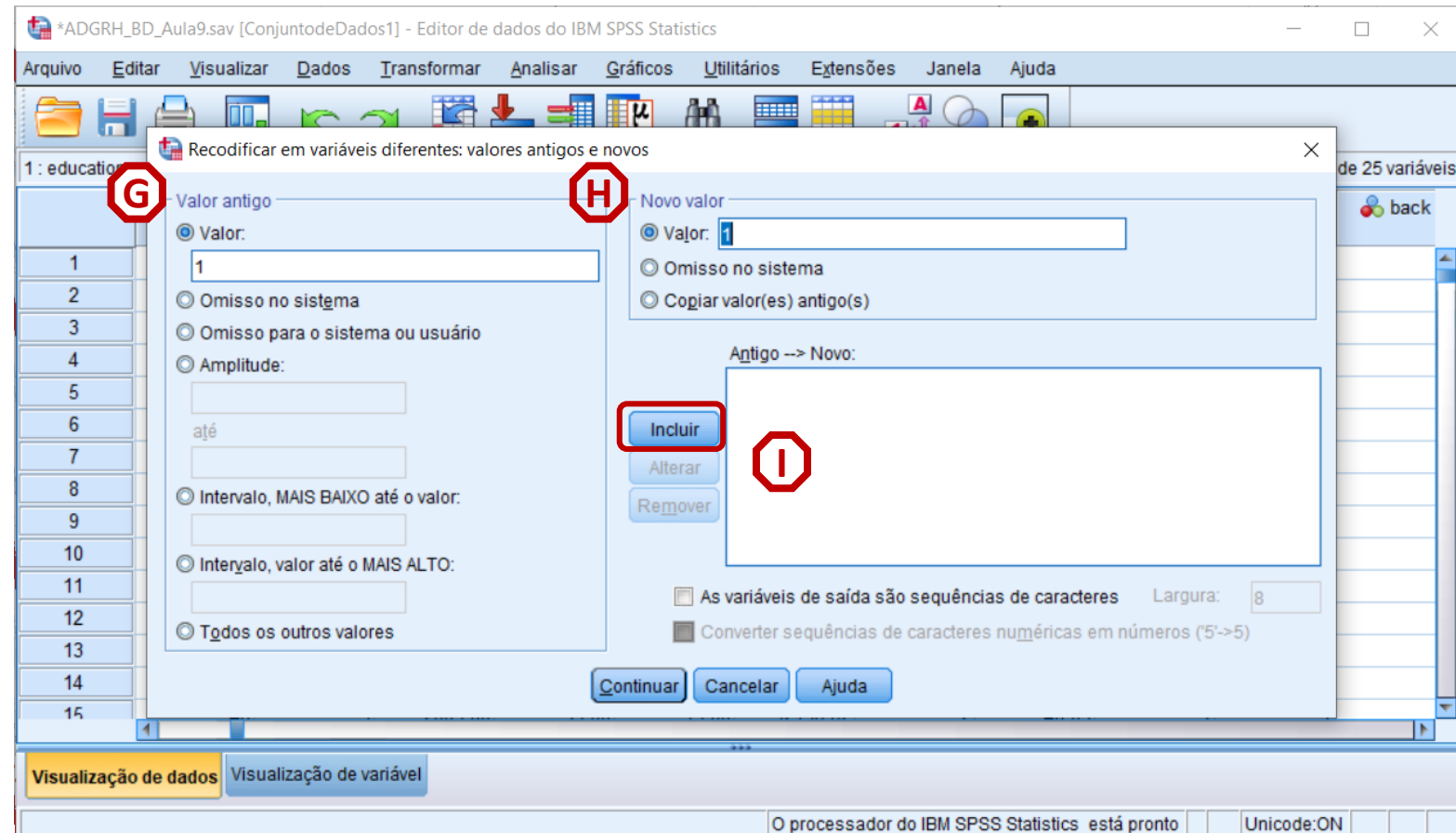
Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON



### Adicionar IVs nominais

- Vamos definir que ao valor '1' na variável 'education'
- Corresponde o valor '1', na nova variável ('educ1')
- Selecionar 'Incluir'

\*ADGRH\_BD\_Aula9.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Recodificar em variáveis diferentes: valores antigos e novos

Valor antigo:
 

- Valor: 1
- 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: 1
- Omisso no sistema
- Copiar valor(es) antigo(s)

Antigo -> Novo:
 

- Incluir** (highlighted)
- Alterar
- Remover

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

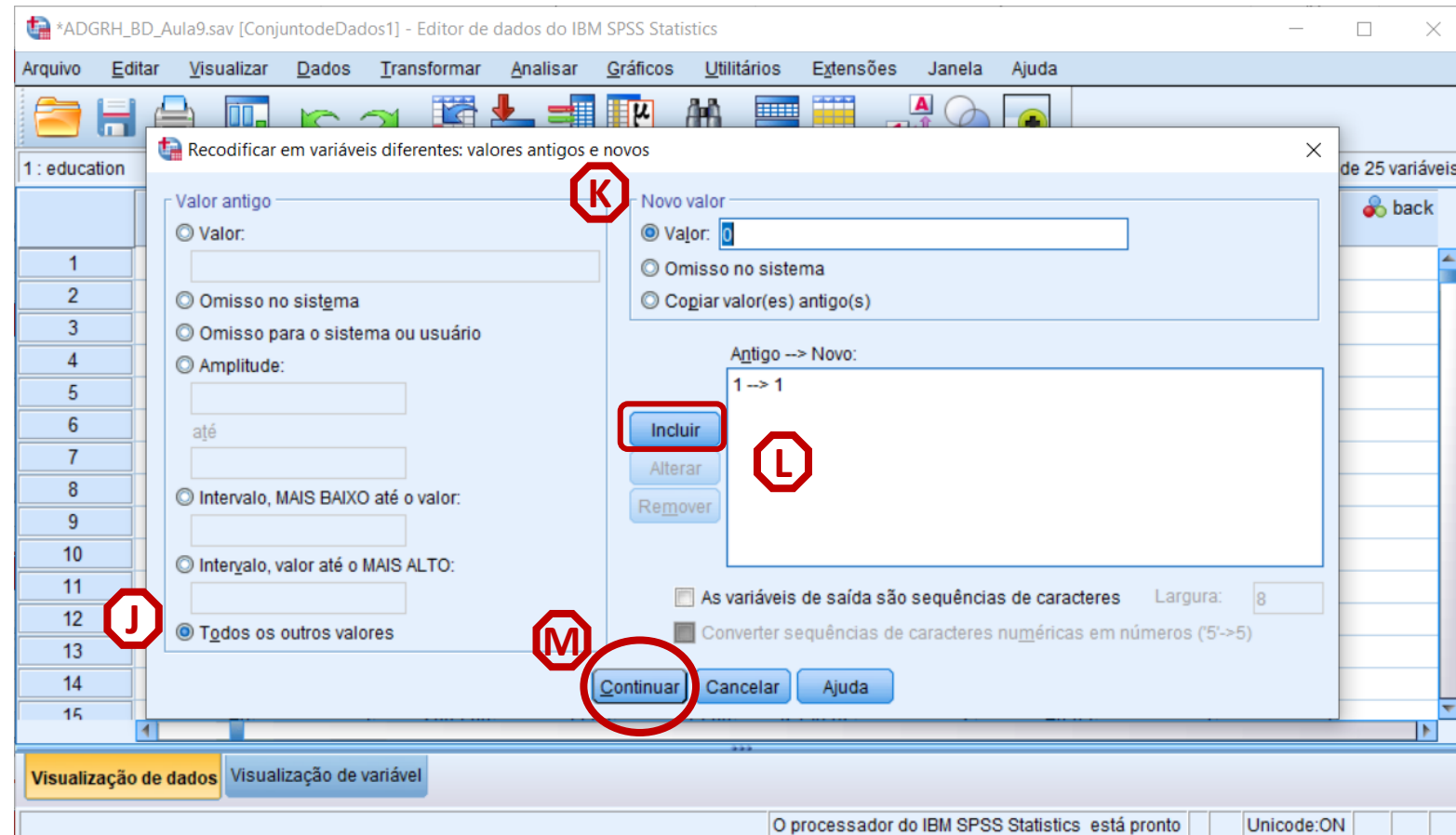
Continuar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

### Adicionar IVs nominais

- Vamos definir que ao valor '1' na variável 'education' G
- Corresponde o valor '1', na nova variável ('educ1') H
- Selecionar 'Incluir' I
- Vamos definir que 'Todos os outros valores' na variável 'education' J
- Corresponde o valor '0', na nova variável ('educ1') K
- Selecionar 'Incluir' L
- Selecionar 'Continuar' / 'OK' M



## Adicionar IVs nominais

- Vamos verificar se a recodificação foi bem sucedida
- Exercício:** repetir o exercício para criar variáveis:  
 educ2  
 educ3  
 educ4  
 educ5

\*ADGRH\_BD\_Aula9.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : educ1 .0 Visível: 26 de 26 variáveis

	fatigue	absent_nr	workcond_sat	evaluation	y_wage2	evaluation_2	sex2	educ1	var	var	var
1	1	9	3	49.08	17529.60	55.17	.00	.00			
2	1	10	2	52.58	24843.10	51.75	.00	.00			
3	1	15	2	53.80	29874.72	54.79	.00	.00			
4	1	10	2	50.42	20994.46	58.30	.00	1.00			
5	1	11	4	48.43	10044.01	49.89	1.00	1.00			
6	1	10	2	54.34	32321.53	52.74	.00	.00			
7	1	6	3	49.65	18757.44	47.60	1.00	.00			
8	1	10	4	46.84	9106.36	52.65	1.00	.00			
9	1	8	2	50.76	20432.69	48.48	1.00	.00			
10	1	17	3	53.33	29327.90	52.40	.00	.00			
11	1	14	2	50.59	18223.72	47.21	.00	1.00			
12	1	7	2	54.64	29668.47	48.39	.00	1.00			
13	1	13	3	56.64	35658.63	50.19	.00	1.00			
14	1	7	3	55.36	35733.01	40.86	1.00	.00			
15	1	11	3	56.09	36160.52	45.36	.00	.00			

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

# ANÁLISE DE DADOS EM GRH

## Aula 9: O modelo de regressão linear

### Adicionar IVs nominais

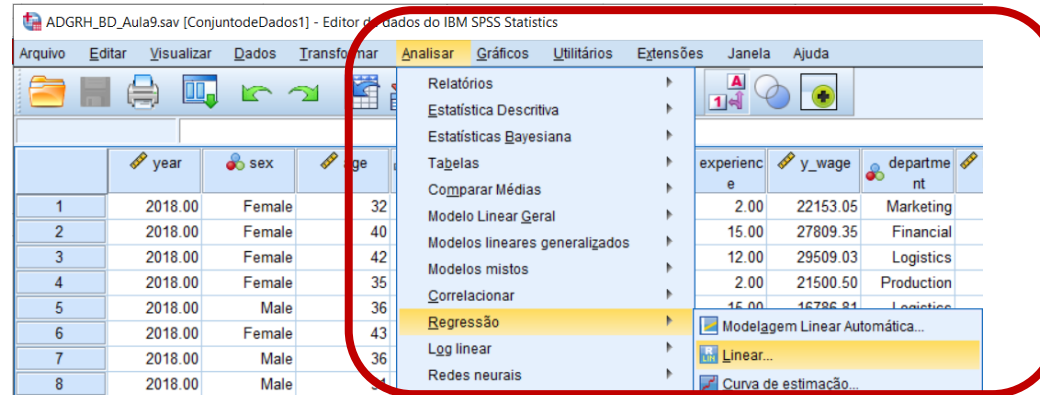
- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Dependente'

**Exercício:** Colocar as variáveis 'sex2', 'experience' e 'evaluation' na caixa 'Independente(s)'

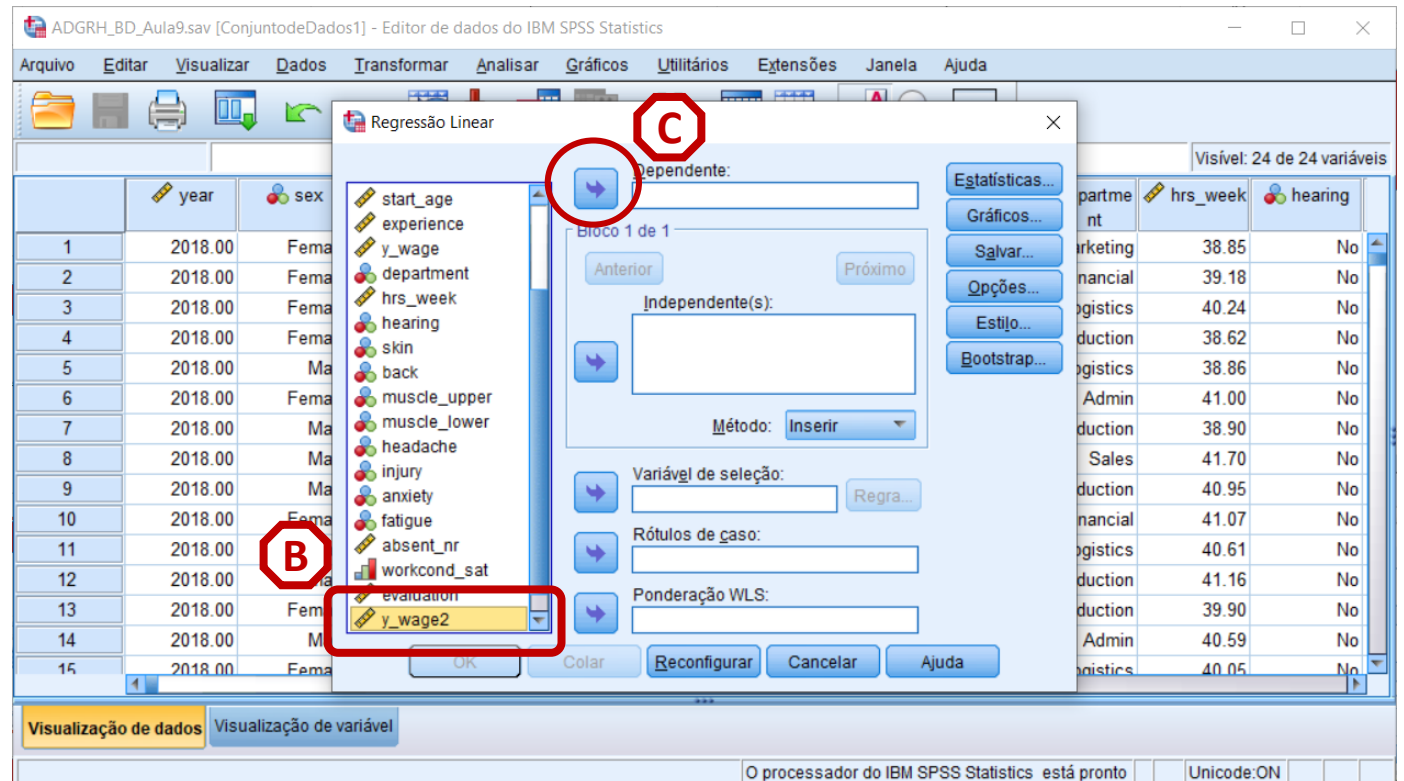
A

B

C



A



C

B

### Adicionar IVs nominais

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
  - Selecionar a variável 'y\_wage2'
  - Colocar na caixa 'Dependente'
  - Colocar na caixa 'Independente(s)'
  - Colocar na caixa 'Dependente'
- Exercício:** Colocar as variáveis 'sex2', 'experience' e 'evaluation' na caixa 'Independente(s)'
- Selecionar as variáveis 'educ1' a 'educ5'
  - Colocar na caixa 'Dependente'

A

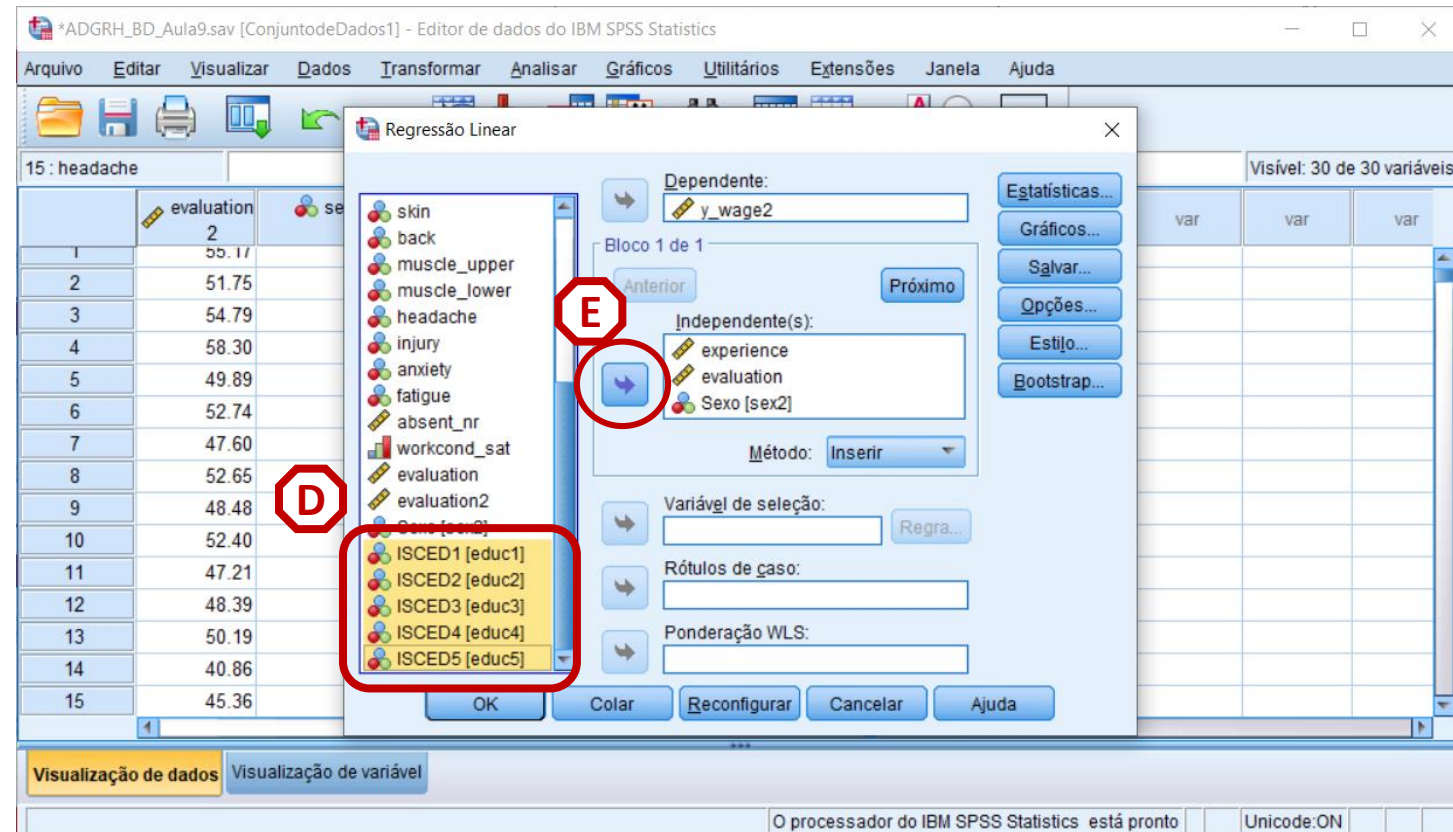
B

C

D

D

E



## Adicionar IVs nominais

- Selecionar 'Analisar' / 'Regressão' / 'Linear'

A

- Selecionar a variável 'y\_wage2'

B

- Colocar na caixa 'Dependente'

C

Exercício: Colocar as variáveis 'sex2', 'experience' e 'evaluation' na caixa 'Independente(s)'

- Selecionar as variáveis 'educ1' a 'educ5'

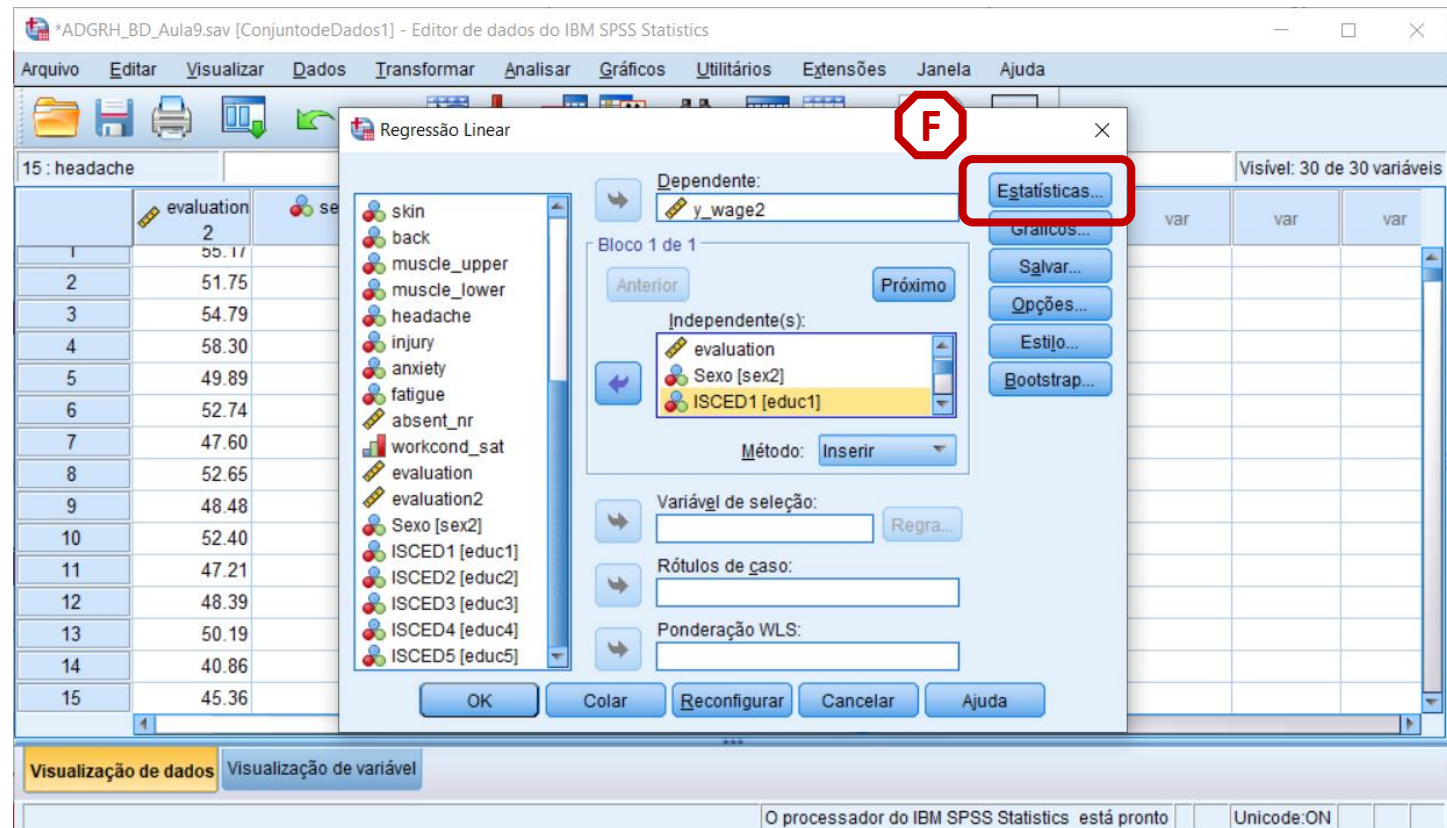
D

- Colocar na caixa 'Dependente'

E

- Selecionar botão 'Estatísticas'

F



\*ADGRH\_BD\_Aula9.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Regressão Linear

Dependente: y\_wage2

Bloco 1 de 1

Independente(s): evaluation, Sexo [sex2], ISCED1 [educ1]

Método: Inserir

Variável de seleção: Regra...

Rótulos de caso:

Ponderação WLS:

OK Colar Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Adicionar IVs nominais

- Selecionar 'Estimativas' **G**
- Selecionar 'Ajuste do modelo' **H**
- Selecionar 'Continuar'/'OK' **I**

The screenshot shows the 'Regressão Linear' dialog box in IBM SPSS Statistics. The 'Dependente:' field is empty. The 'Coeficientes de regressão' section has 'Estimativas' checked (marked with a red circle and 'G'). The 'Residuais' section has 'Ajuste do modelo' checked (marked with a red circle and 'H'). The 'Continuar' button is circled in red (marked with a red circle and 'I'). The background shows a data table with columns like 'year', 'sex', 'hrs\_week', and 'hearing'.

	year	sex	hrs_week	hearing
1	2018.00	Fema		
2	2018.00	Fema		
3	2018.00	Fema		
4	2018.00	Fema		
5	2018.00	Ma		
6	2018.00	Fema		
7	2018.00	Ma		
8	2018.00	Ma		
9	2018.00	Ma		
10	2018.00	Fema		
11	2018.00	Fema		
12	2018.00	Fema		
13	2018.00	Fema		
14	2018.00	Ma		
15	2018.00	Fema		

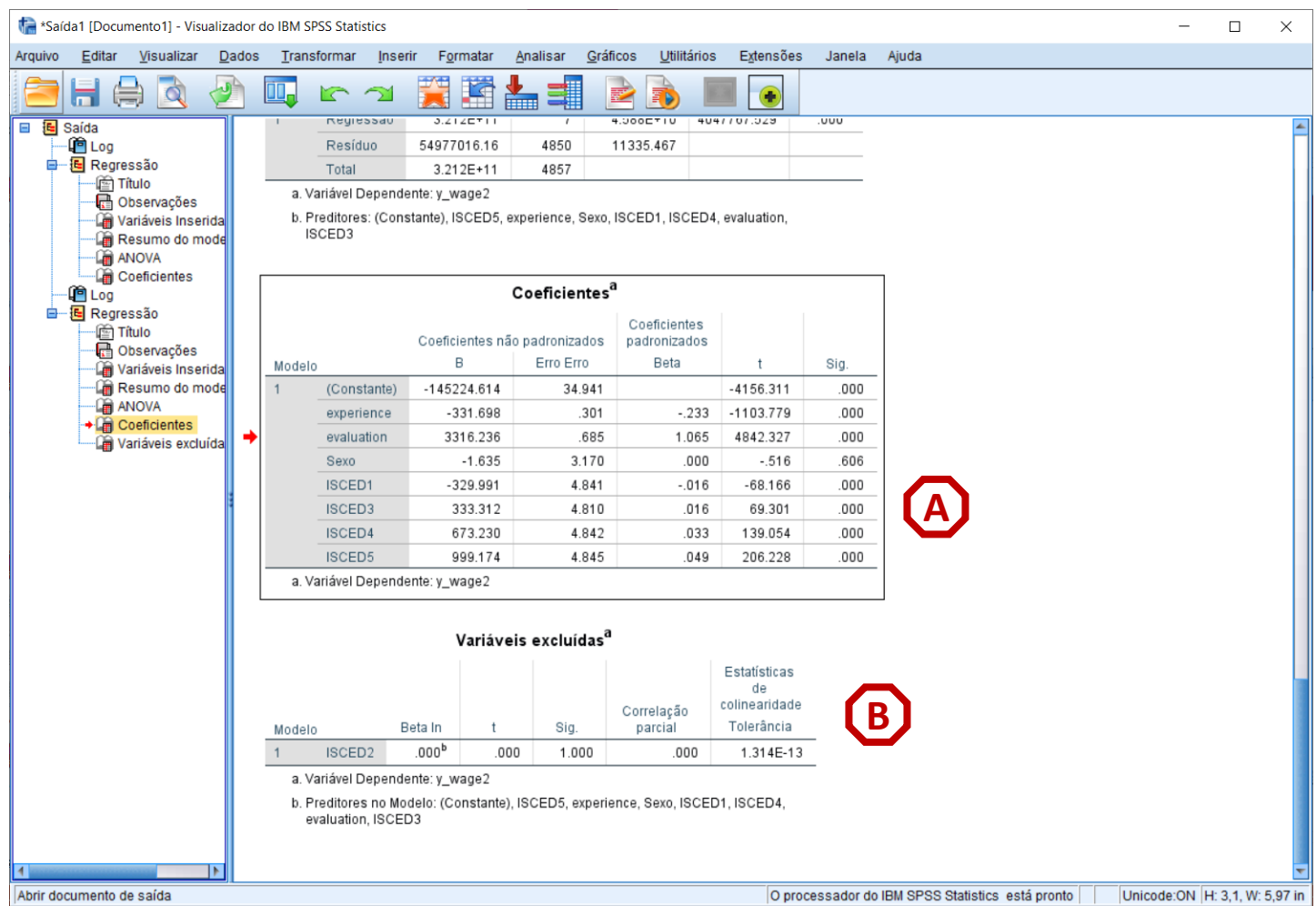
### Adicionar IVs nominais

- O que diz a tabela sobre o efeito das variáveis independentes?

O valor Sig. é menor que 0.01

→ A variáveis de educação têm uma relação estatisticamente significativa com a DV, com um grau de confiança a 99%

A exceção é a variável 'ISCED2', que foi excluída do modelo por ter sido identificada a possibilidade de colinearidade com alguma das outras IVs

Resumo da regressão

	Resíduo	Total
Resíduo	54977016.16	4850
Total	3.212E+11	4857

a. Variável Dependente: y\_wage2  
b. Preditores: (Constante), ISCED5, experience, Sexo, ISCED1, ISCED4, evaluation, ISCED3

**Coeficientes<sup>a</sup>**

Modelo		Coeficientes não padronizados		Coeficientes padronizados		Sig.
		B	Erro Erro	Beta	t	
1	(Constante)	-145224.614	34.941		-4156.311	.000
	experience	-.331.698	.301	-.233	-1103.779	.000
	evaluation	3316.236	.685	1.065	4842.327	.000
	Sexo	-1.635	3.170	.000	-.516	.606
	ISCED1	-329.991	4.841	-.016	-68.166	.000
	ISCED3	333.312	4.810	.016	69.301	.000
	ISCED4	673.230	4.842	.033	139.054	.000
	ISCED5	999.174	4.845	.049	206.228	.000

a. Variável Dependente: y\_wage2

**Variáveis excluídas<sup>a</sup>**

Modelo		Beta In	t	Sig.	Correlação parcial	Estatísticas de colinearidade	
						Tolerância	
1	ISCED2	.000 <sup>b</sup>	.000	1.000	.000	1.314E-13	

a. Variável Dependente: y\_wage2  
b. Preditores no Modelo: (Constante), ISCED5, experience, Sexo, ISCED1, ISCED4, evaluation, ISCED3





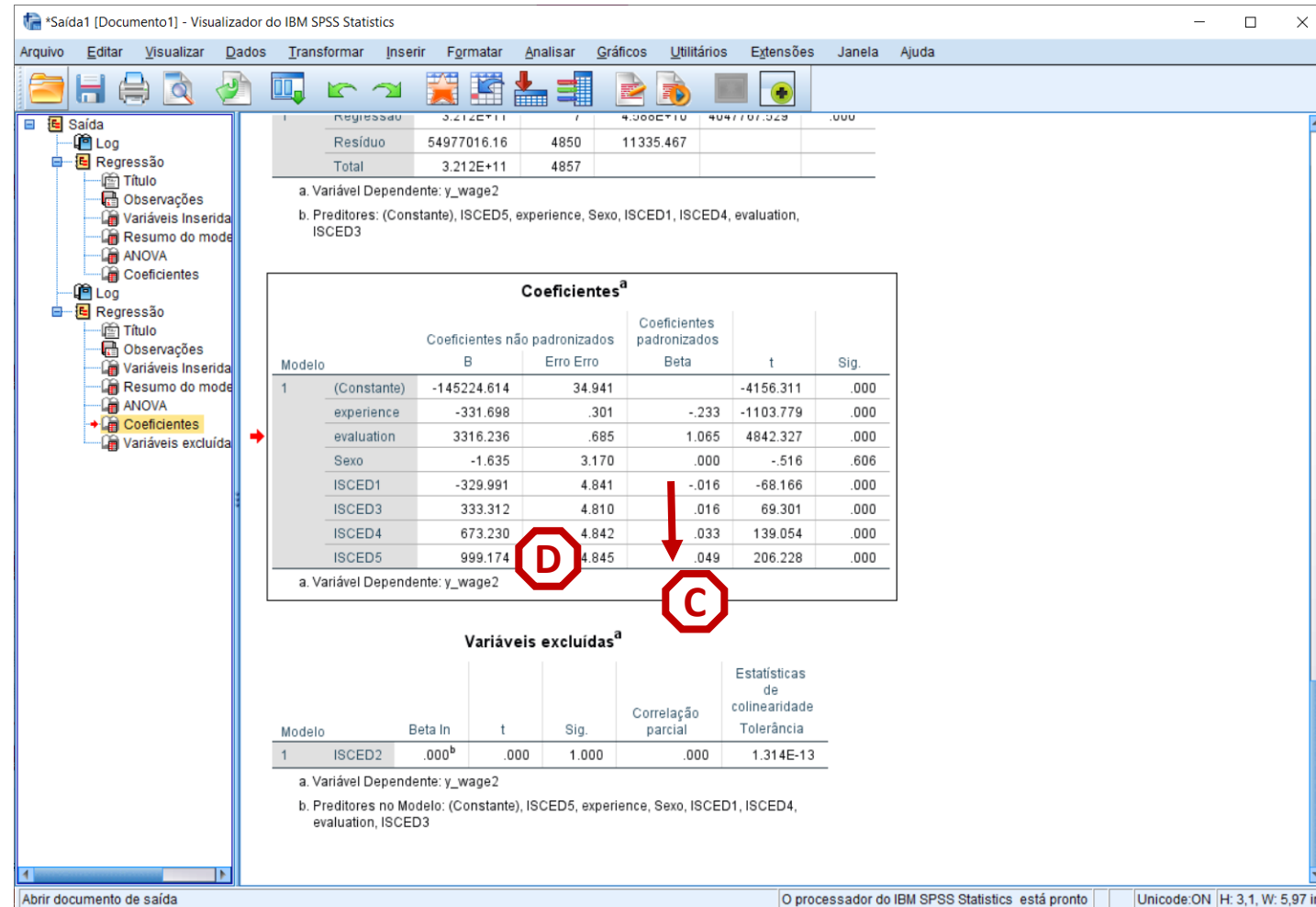
### Adicionar IVs nominais

- O que diz a tabela sobre o efeito das variáveis independentes?

À medida que aumenta o grau de educação aumenta o efeito sobre o salário anual.



Ter uma licenciatura aumenta o salário (anual) em 999 Euros, em relação a todos os que não têm uma licenciatura.

Visualizador do IBM SPSS Statistics

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

Resíduo	3.212E+11	7	4.500E+10	4047107.323	.000
Resíduo	54977016.16	4850	11335.467		
Total	3.212E+11	4857			

a. Variável Dependente: y\_wage2  
b. Preditores: (Constante), ISCED5, experience, Sexo, ISCED1, ISCED4, evaluation, ISCED3

Modelo		Coeficientes não padronizados		Coeficientes padronizados		
		B	Erro Erro	Beta	t	Sig.
1	(Constante)	-145224.614	34.941		-4156.311	.000
	experience	-.331.698	.301	-.233	-1103.779	.000
	evaluation	3316.236	.685	1.065	4842.327	.000
	Sexo	-1.635	3.170	.000	-.516	.606
	ISCED1	-.329.991	4.841	-.016	-68.166	.000
	ISCED3	333.312	4.810	.016	69.301	.000
	ISCED4	673.230	4.842	.033	139.054	.000
	ISCED5	999.174	4.845	.049	206.228	.000

a. Variável Dependente: y\_wage2

Modelo	Beta In	t	Sig.	Correlação parcial	Estadísticas de colinearidade Tolerância	
1	ISCED2	.000 <sup>b</sup>	.000	1.000	.000	1.314E-13

a. Variável Dependente: y\_wage2  
b. Preditores no Modelo: (Constante), ISCED5, experience, Sexo, ISCED1, ISCED4, evaluation, ISCED3



# Implementação do Modelo de Regressão Linear

*5. Qual é o modelo mais eficiente? (I)*

## Escolha do modelo mais eficiente

- **Objectivo:**
  - Estimar um modelo de regressão que tenha a maior capacidade explicativa possível, mas que seja o mais parsimonioso possível



## Escolha do modelo mais eficiente

- O SPSS oferece uma série de opções para se identificar o modelo mais eficiente



\*ADGRH\_BD\_Aula9.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

15 : headache

	evaluation	y_wage2
1	55.17	
2	51.75	
3	54.79	
4	58.30	
5	49.89	
6	52.74	
7	47.60	
8	52.65	
9	48.48	
10	52.40	
11	47.21	
12	48.39	
13	50.19	
14	40.86	
15	45.36	

Regressão Linear

Dependente: y\_wage2

Bloco 1 de 1

Independente(s): experience, evaluation, Sexo [sex2]

Método: Inserir

Variável de seleção:

Rótulos de caso:

Ponderação WLS:

OK Colar Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

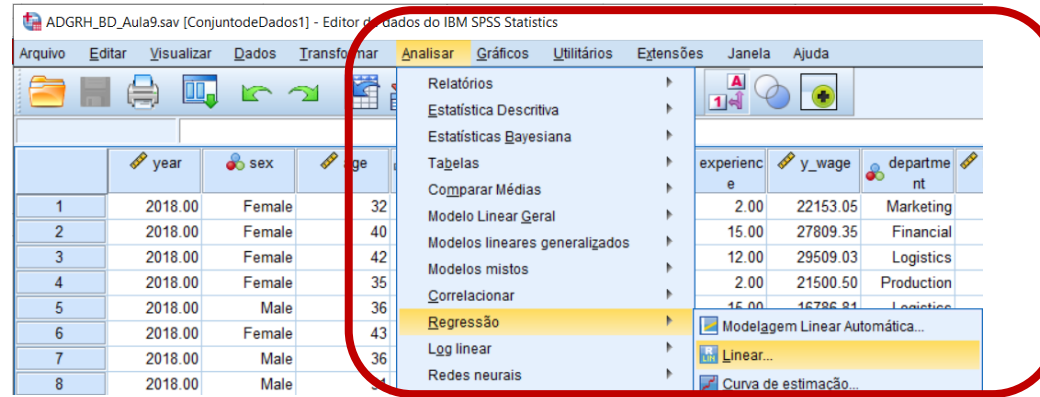
O processador do IBM SPSS Statistics está pronto Unicode:ON

Método	Definição
Inserir	<ul style="list-style-type: none"> <li>Todas as variáveis são inseridas ao mesmo tempo</li> </ul>
Remover	<ul style="list-style-type: none"> <li>Todas as variáveis são removidas num único passo.</li> </ul>
Backward (Backward elimination)	<ul style="list-style-type: none"> <li>Procedimento de seleção em que todas as variáveis são inseridas na equação e removidas sequencialmente.</li> <li>A variável com a menor correlação parcial com o a variável dependente é a primeira a ser considerada para remoção. Se atender ao critério de eliminação, é removida. E assim consecutivamente.</li> <li>O procedimento pára quando não há variáveis na equação que satisfaçam os critérios de remoção.</li> </ul>
Redireccionar (forward selection) 😊	<ul style="list-style-type: none"> <li>Procedimento de seleção no qual as variáveis são inseridas sequencialmente no modelo.</li> <li>A primeira variável considerada para entrada na equação é aquela com a maior correlação positiva ou negativa com a variável dependente. Esta variável é inserida na equação apenas se satisfizer o critério de entrada. E assim consecutivamente.</li> <li>O procedimento pára quando houver variáveis que atendam ao critério de entrada.</li> </ul>
Stepwise 😊	<ul style="list-style-type: none"> <li>Em cada etapa, a variável independente que não está na equação que tem a menor probabilidade de F é inserida, se essa probabilidade for suficientemente pequena.</li> <li>Variáveis já na equação de regressão são removidos se sua probabilidade de F se tornar suficientemente grande.</li> <li>O método pára quando não houver mais variáveis elegíveis para inclusão ou remoção.</li> </ul>

### Escolha do modelo mais eficiente

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Dependente'

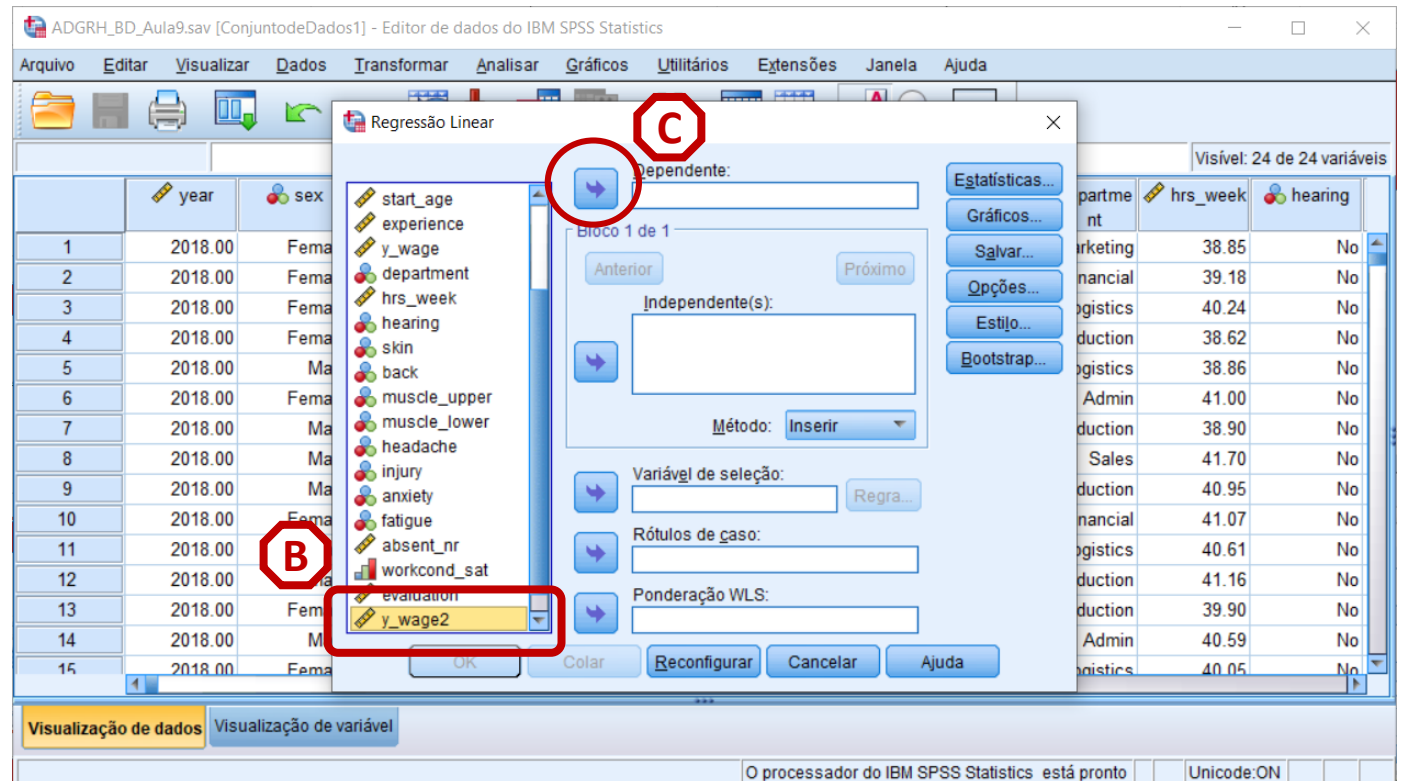
**Exercício:** Colocar as variáveis 'sex2', 'experience' e 'evaluation' na caixa 'Independente(s)'



A

B

C



B

C

### Escolha do modelo mais eficiente

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Dependente'

Exercício: Colocar as variáveis 'sex2', 'experience' e 'evaluation' na caixa 'Independente(s)'

- Selecionar o Método 'Redirecionar'
- Selecionar botão 'Estatísticas'

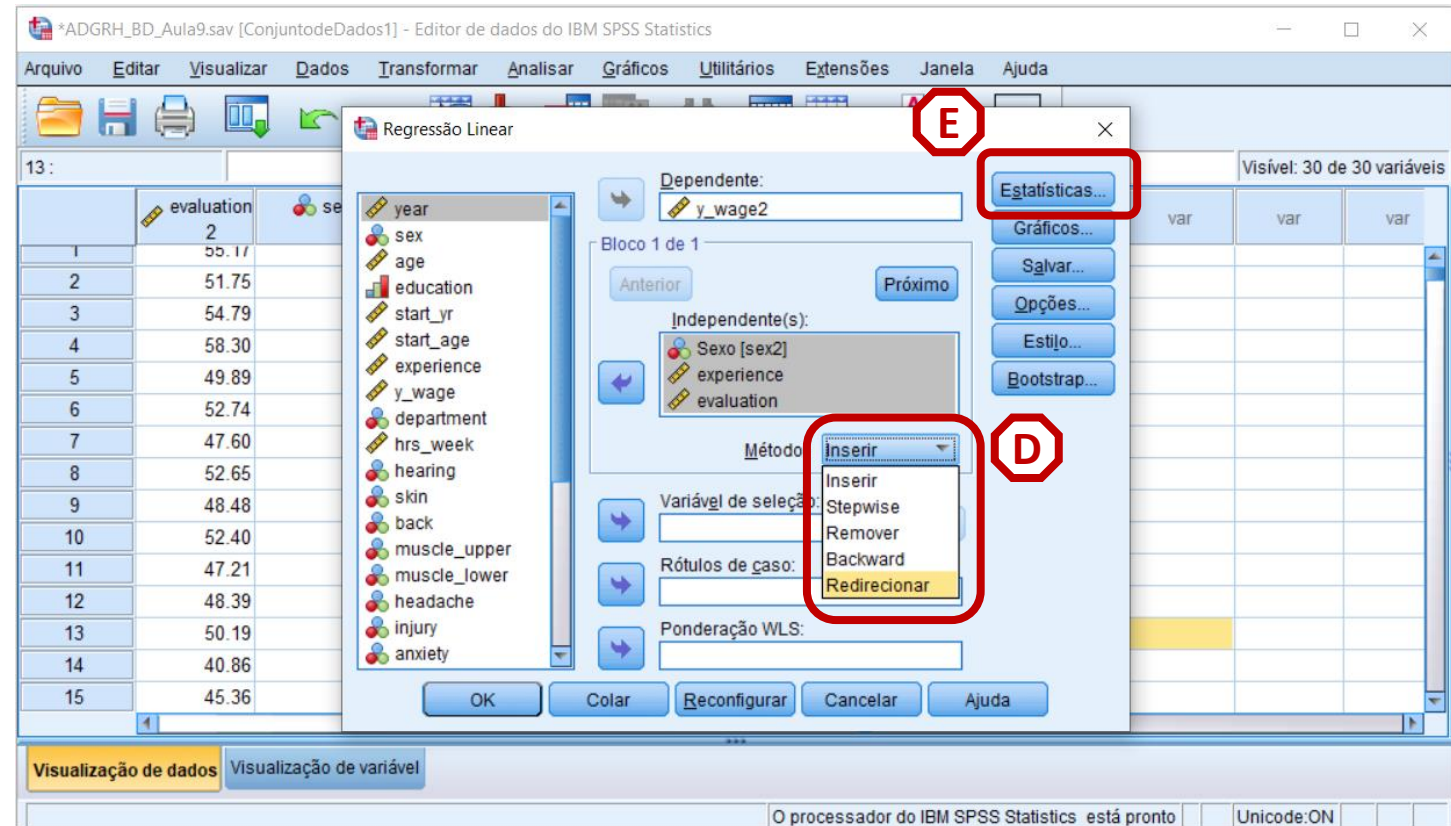
A

B

C

D

E



## Escolha do modelo mais eficiente

- Selecionar 'Estimativas'
- Selecionar 'Ajuste do modelo'
- Selecionar 'Continuar'/'OK'

F

G

H

ADGRH\_BD\_Aula9.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Regressão Linear

Dependente:

Regressão Linear: estatísticas

**F**  Estimativas  
 Intervalos de confiança  
Nível (%): 95  
 Matriz de covariâncias

**G**  Ajuste do modelo  
 Alteração de R quadrado  
 Descritivos  
 Correlações parciais e de parte  
 Diagnósticos de colinearidade

Residuais

Durbin-Watson  
 Diagnóstico por caso  
 Valgres discrepantes r de fora: 3 desvios padrão  
 Todos os casos

**H**

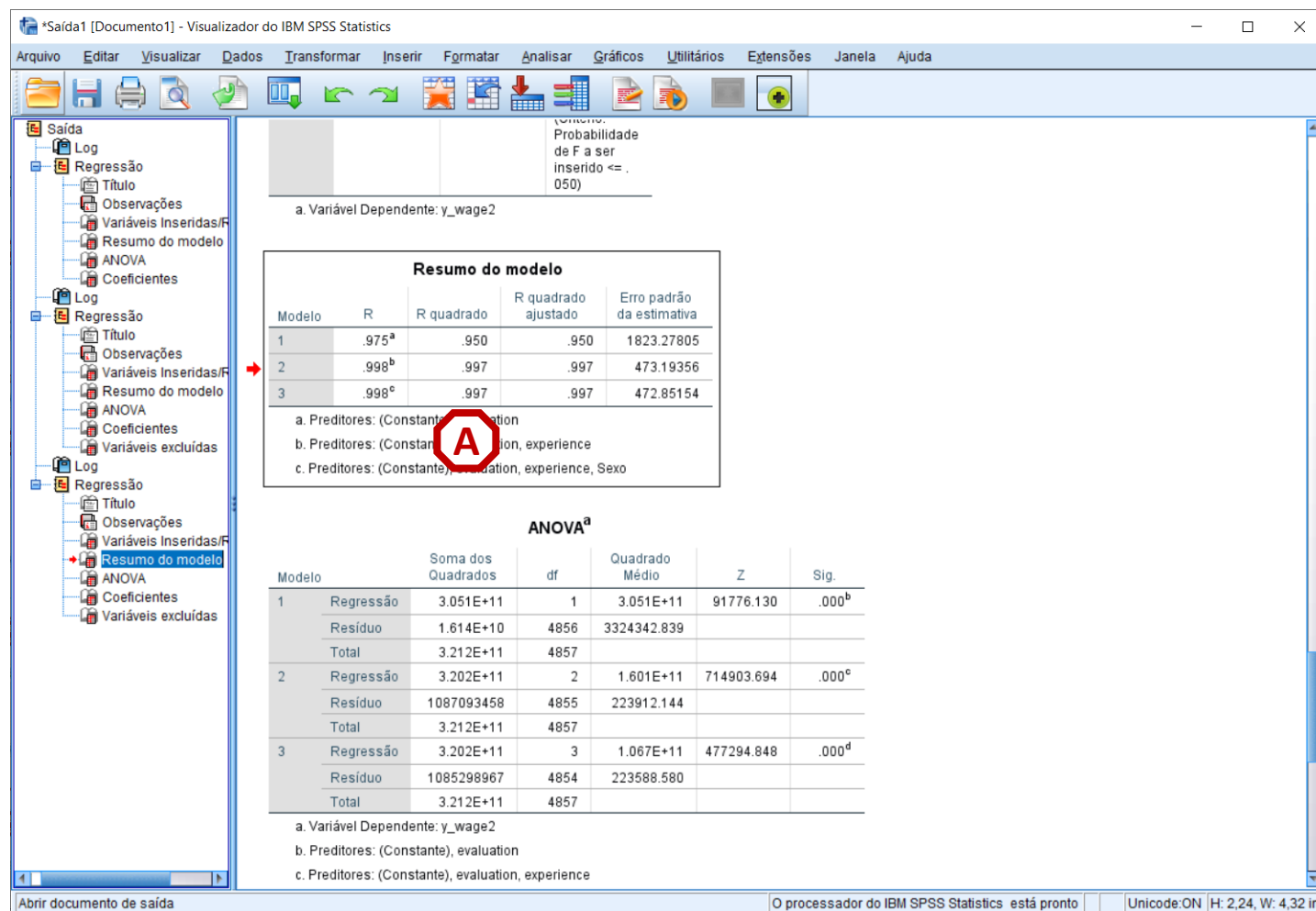
Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON



## Escolha do modelo mais eficiente

- O  $R^2$  da Tabela de Resumo do modelo diz-nos que adicionar 'sex2' ao nosso modelo não aumenta o poder explicativo do modelo

a. Variável Dependente: y\_wage2

Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa
1	.975 <sup>a</sup>	.950	.950	1823.27805
2	.998 <sup>b</sup>	.997	.997	473.19356
3	.998 <sup>c</sup>	.997	.997	472.85154

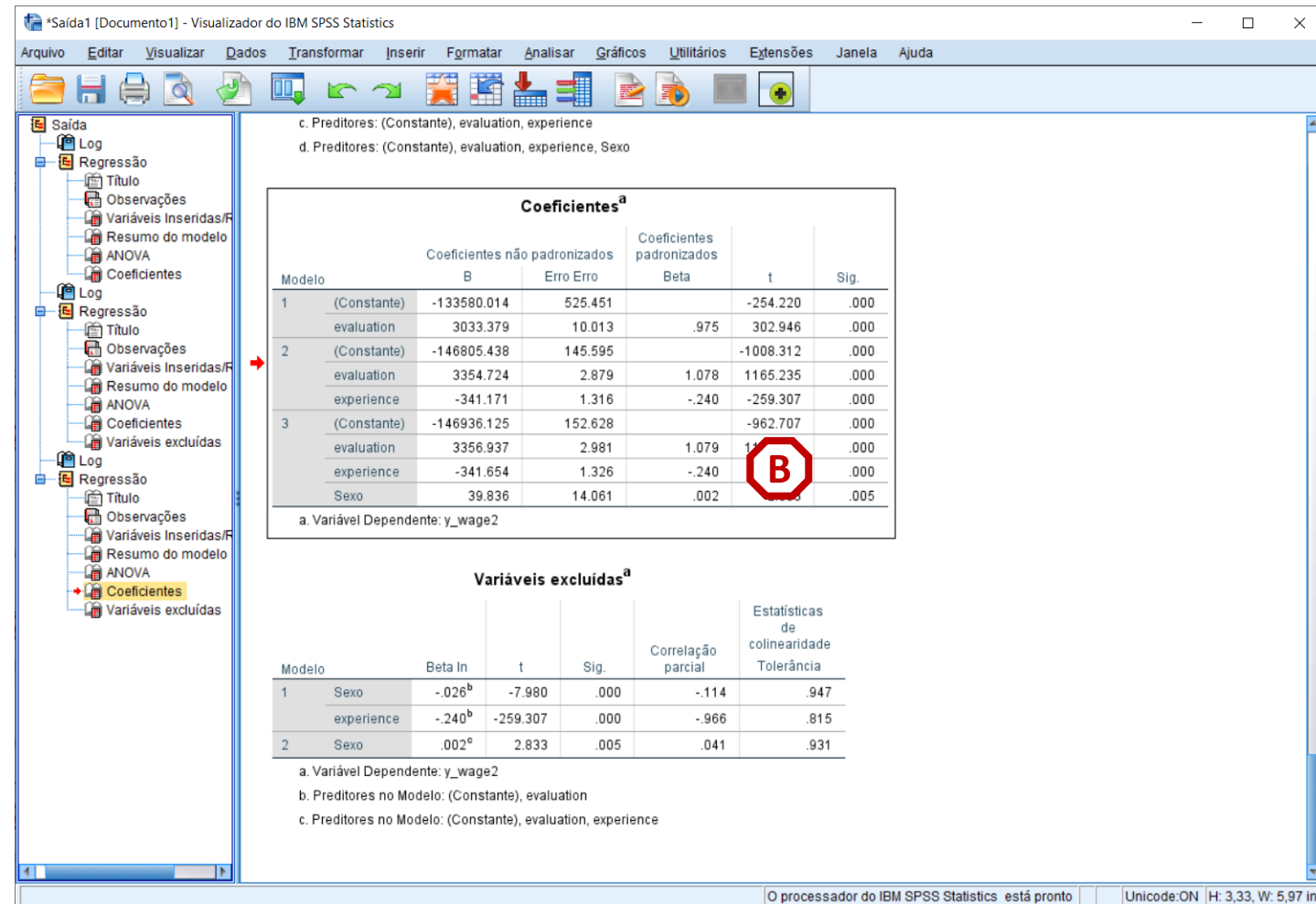
a. Preditores: (Constante), evaluation  
 b. Preditores: (Constante), evaluation, experience  
 c. Preditores: (Constante), evaluation, experience, Sexo

Modelo		Soma dos Quadrados	df	Quadrado Médio	Z	Sig.
1	Regressão	3.051E+11	1	3.051E+11	91776.130	.000 <sup>b</sup>
	Resíduo	1.614E+10	4856	3324342.839		
	Total	3.212E+11	4857			
2	Regressão	3.202E+11	2	1.601E+11	714903.694	.000 <sup>c</sup>
	Resíduo	1087093458	4855	223912.144		
	Total	3.212E+11	4857			
3	Regressão	3.202E+11	3	1.067E+11	477294.848	.000 <sup>d</sup>
	Resíduo	1085298967	4854	223588.580		
	Total	3.212E+11	4857			

a. Variável Dependente: y\_wage2  
 b. Preditores: (Constante), evaluation  
 c. Preditores: (Constante), evaluation, experience

### Escolha do modelo mais eficiente

- A Tabela dos Coeficientes (ver Coeficientes Padronizados Beta), sugere que adicionar ‘experiência’ ao nosso modelo tem um impacto significativo sobre a variável ‘avaliação’.
- Adicionar a variável ‘sexo’ não tem influência sobre as outras variáveis

c. Preditores: (Constante), evaluation, experience  
 d. Preditores: (Constante), evaluation, experience, Sexo

Modelo		Coeficientes não padronizados		Coeficientes padronizados		
		B	Erro Erro	Beta	t	Sig.
1	(Constante)	-133580.014	525.451		-254.220	.000
	evaluation	3033.379	10.013	.975	302.946	.000
2	(Constante)	-146805.438	145.595		-1008.312	.000
	evaluation	3354.724	2.879	1.078	1165.235	.000
	experience	-341.171	1.316	-.240	-259.307	.000
3	(Constante)	-146936.125	152.628		-962.707	.000
	evaluation	3356.937	2.981	1.079	1165.235	.000
	experience	-341.654	1.326	-.240	-259.307	.000
	Sexo	39.836	14.061	.002	2.833	.005

a. Variável Dependente: y\_wage2

Modelo		Beta In	t	Sig.	Correlação parcial	Estadísticas de colinearidade
						Tolerância
1	Sexo	-.026 <sup>b</sup>	-7.980	.000	-.114	.947
	experience	-.240 <sup>b</sup>	-259.307	.000	-.966	.815
2	Sexo	.002 <sup>c</sup>	2.833	.005	.041	.931

a. Variável Dependente: y\_wage2  
 b. Preditores no Modelo: (Constante), evaluation  
 c. Preditores no Modelo: (Constante), evaluation, experience

# Implementação do Modelo de Regressão Linear

*6. Qual é o modelo mais eficiente (II)? A Regressão Stepwise*

### Regressão Stepwise

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y\_wage2'
- Colocar na caixa 'Dependente'

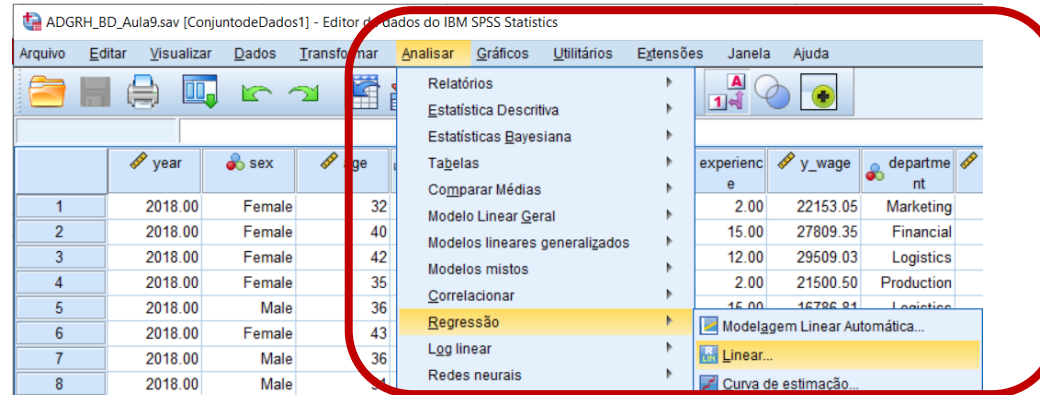
Exercício: Colocar as variáveis 'sex2', 'experience' e 'evaluation' na caixa 'Independente(s)'

- Selecionar o Método 'Redirecionar'
- Selecionar botão 'Estatísticas'

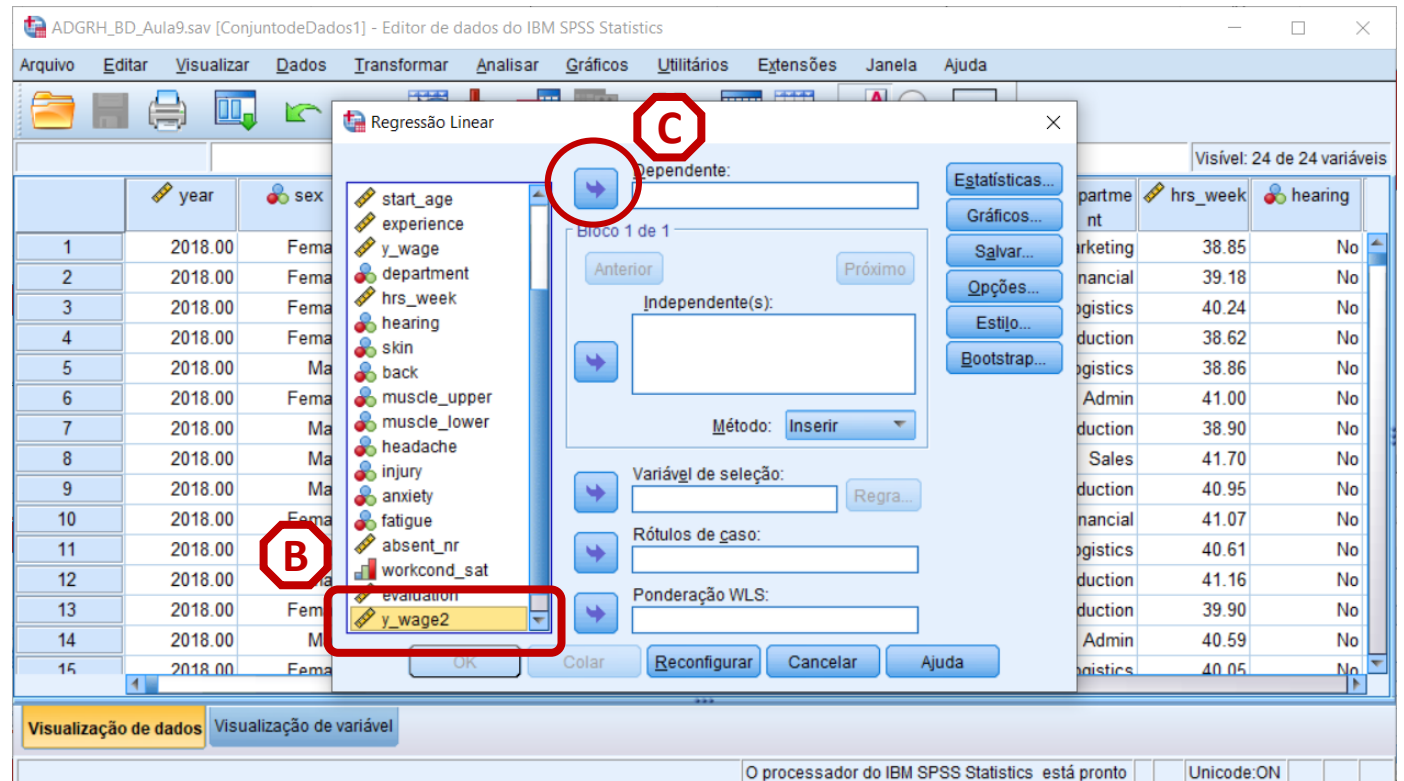
A

B

C



A



B

C

## Regressão Stepwise

- Selecionar 'Analisar' / 'Regressão' / 'Linear'

- Selecionar a variável 'y\_wage2'

- Colocar na caixa 'Dependente'

Exercício: Colocar as variáveis 'sex2', 'experience' e 'evaluation' na caixa 'Independente(s)'

- Selecionar o Método 'Redirecionar'

- Selecionar botão 'Estatísticas'

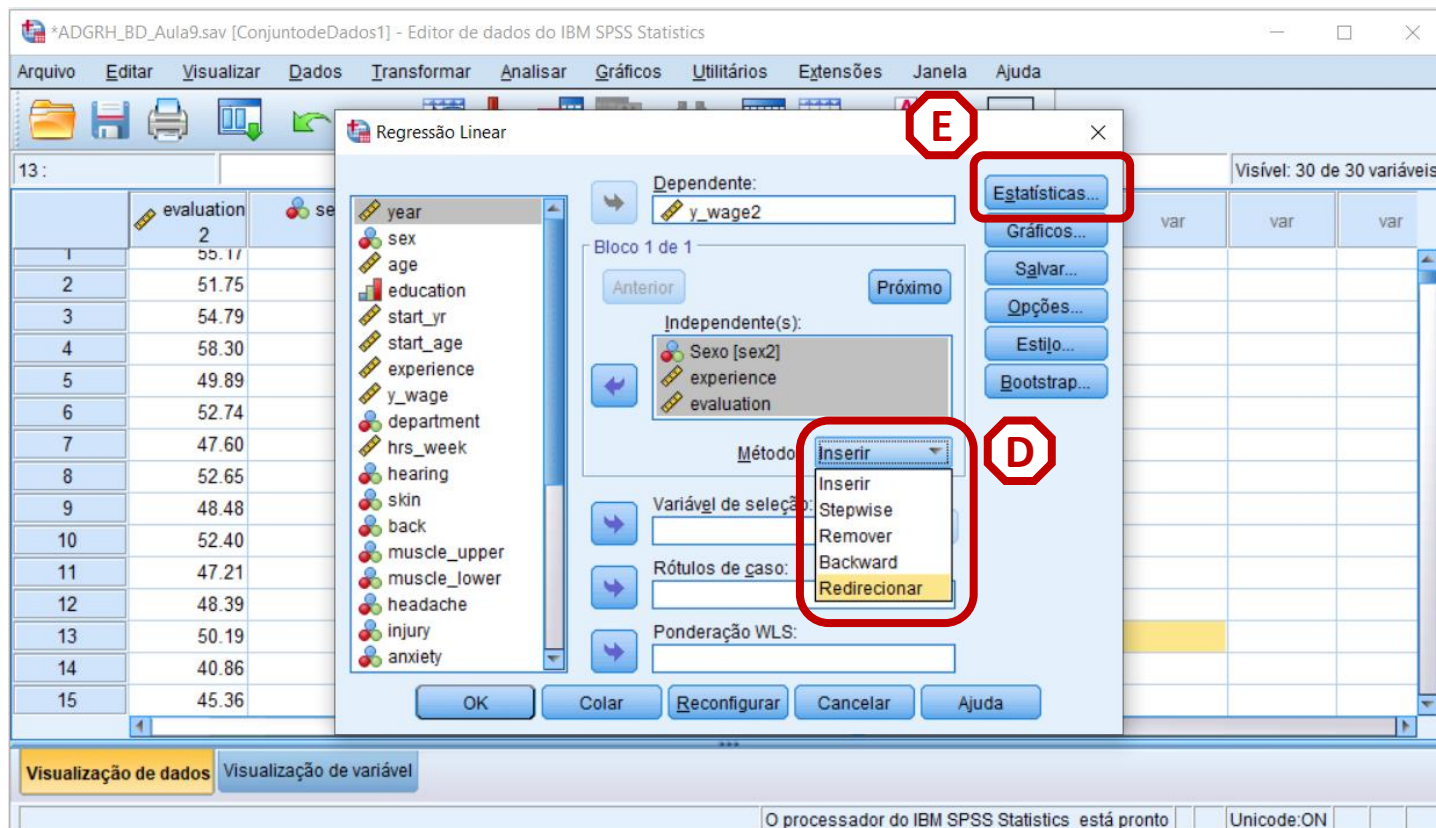
A

B

C

D

E



	evaluation	y_wage2
1	55.17	
2	51.75	
3	54.79	
4	58.30	
5	49.89	
6	52.74	
7	47.60	
8	52.65	
9	48.48	
10	52.40	
11	47.21	
12	48.39	
13	50.19	
14	40.86	
15	45.36	

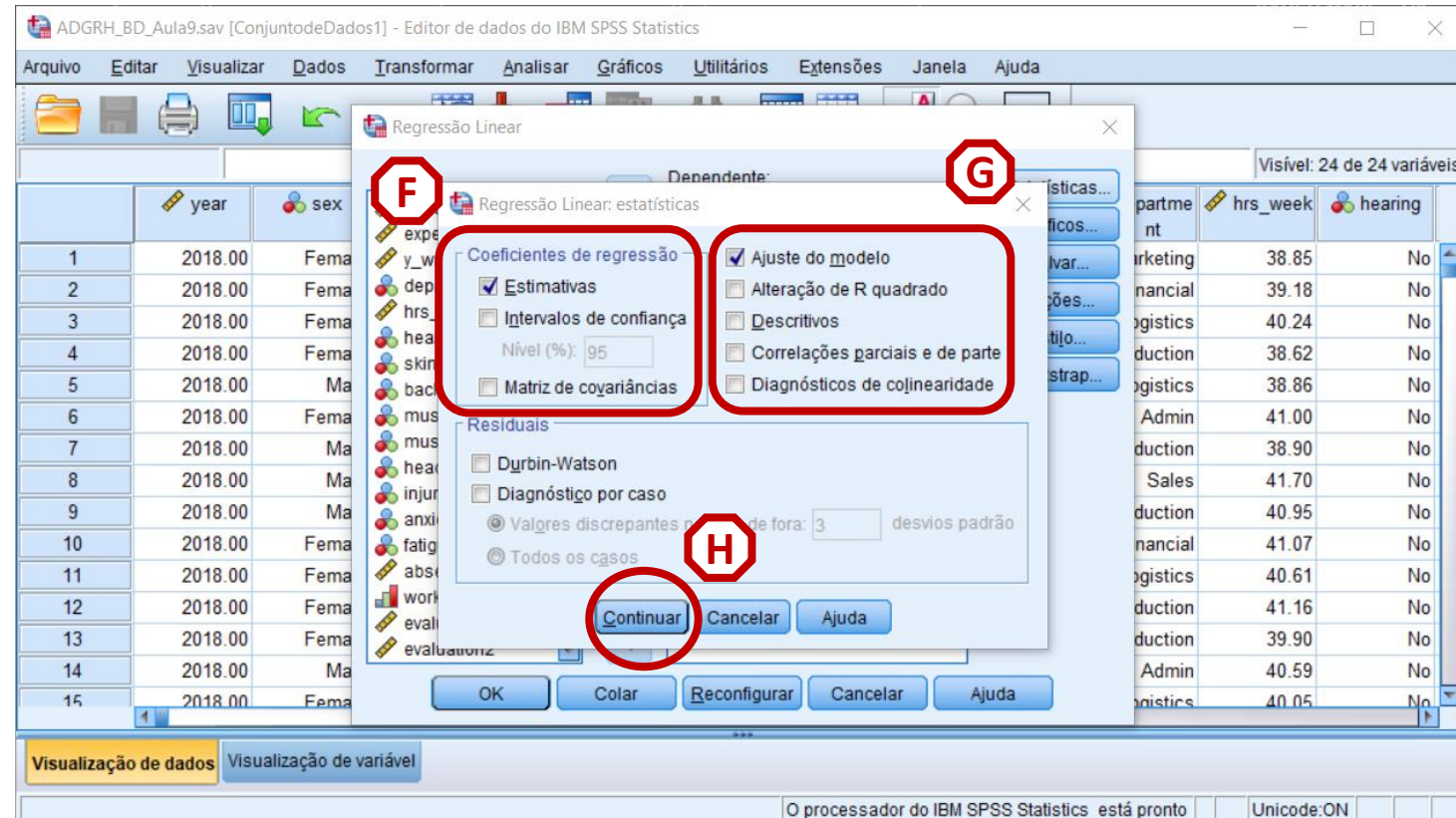
### Regressão Stepwise

- Selecionar 'Estimativas'
- Selecionar 'Ajuste do modelo'
- Selecionar 'Continuar'/'OK'

F

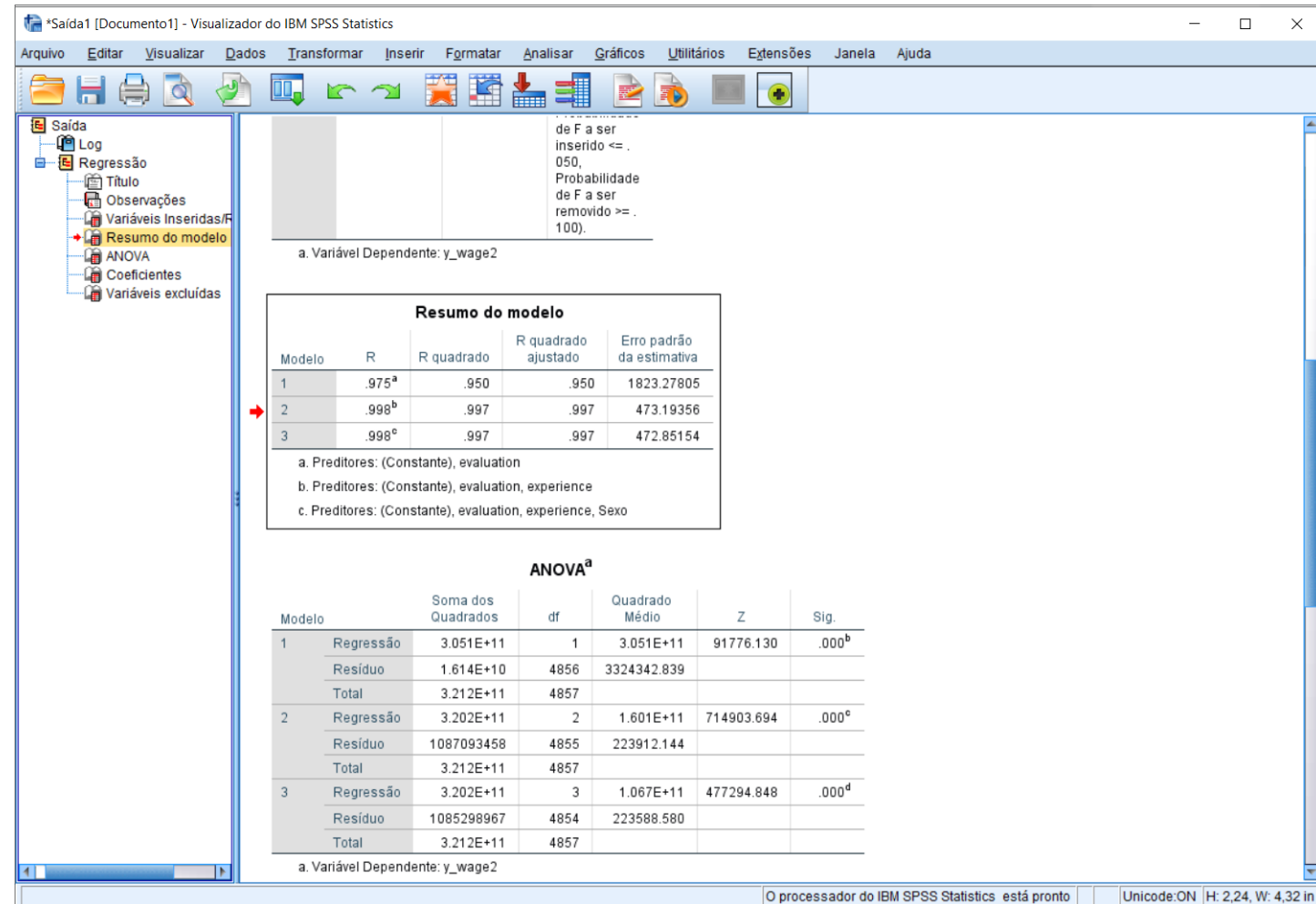
G

H



## Regressão Stepwise

- Os resultados não são diferentes do modelo anterior
- O modelo é demasiado simples...



\*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

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

de F a ser inserido <= .050, Probabilidade de F a ser removido >= .100.

a. Variável Dependente: y\_wage2

Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa
1	.975 <sup>a</sup>	.950	.950	1823.27805
2	.998 <sup>b</sup>	.997	.997	473.19356
3	.998 <sup>c</sup>	.997	.997	472.85154

a. Preditores: (Constante), evaluation  
b. Preditores: (Constante), evaluation, experience  
c. Preditores: (Constante), evaluation, experience, Sexo

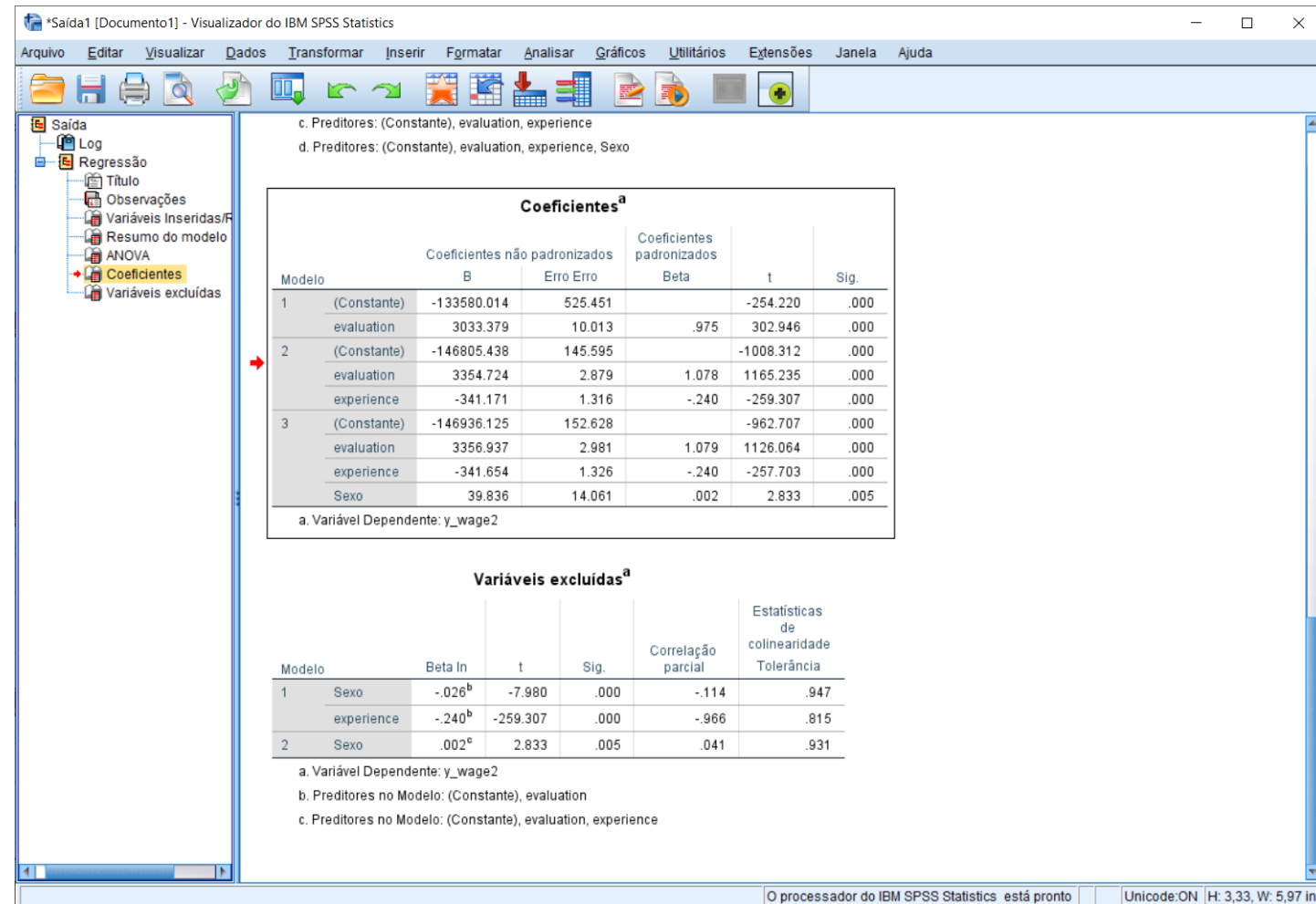
Modelo		Soma dos Quadrados	df	Quadrado Médio	Z	Sig.
1	Regressão	3.051E+11	1	3.051E+11	91776.130	.000 <sup>b</sup>
	Resíduo	1.614E+10	4856	3324342.839		
	Total	3.212E+11	4857			
2	Regressão	3.202E+11	2	1.601E+11	714903.694	.000 <sup>c</sup>
	Resíduo	1087093458	4855	223912.144		
	Total	3.212E+11	4857			
3	Regressão	3.202E+11	3	1.067E+11	477294.848	.000 <sup>d</sup>
	Resíduo	1085298967	4854	223588.580		
	Total	3.212E+11	4857			

a. Variável Dependente: y\_wage2

O processador do IBM SPSS Statistics está pronto | Unicode:ON | H: 2,24, W: 4,32 in

### Regressão Stepwise

- Os resultados não são diferentes do modelo anterior
- O modelo é demasiado simples...



\*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

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

Saída  
Log  
Regressão  
Título  
Observações  
Variáveis inseridas/...  
Resumo do modelo  
ANOVA  
Coeficientes  
Variáveis excluídas

c. Preditores: (Constante), evaluation, experience  
d. Preditores: (Constante), evaluation, experience, Sexo

#### Coeficientes<sup>a</sup>

Modelo		Coeficientes não padronizados		Coeficientes padronizados		
		B	Erro Erro	Beta	t	Sig.
1	(Constante)	-133580.014	525.451		-254.220	.000
	evaluation	3033.379	10.013	.975	302.946	.000
2	(Constante)	-146805.438	145.595		-1008.312	.000
	evaluation	3354.724	2.879	1.078	1165.235	.000
	experience	-341.171	1.316	-.240	-259.307	.000
3	(Constante)	-146936.125	152.628		-962.707	.000
	evaluation	3356.937	2.981	1.079	1126.064	.000
	experience	-341.654	1.326	-.240	-257.703	.000
	Sexo	39.836	14.061	.002	2.833	.005

a. Variável Dependente: y\_wage2

#### Variáveis excluídas<sup>a</sup>

Modelo		Beta In	t	Sig.	Correlação parcial	Estatísticas de colinearidade Tolerância
1	Sexo	-.026 <sup>b</sup>	-7.980	.000	-.114	.947
	experience	-.240 <sup>b</sup>	-259.307	.000	-.966	.815
2	Sexo	.002 <sup>c</sup>	2.833	.005	.041	.931

a. Variável Dependente: y\_wage2  
b. Preditores no Modelo: (Constante), evaluation  
c. Preditores no Modelo: (Constante), evaluation, experience

O processador do IBM SPSS Statistics está pronto | Unicode:ON | H: 3,33, W: 5,97 in





LISBON  
SCHOOL OF  
ECONOMICS &  
MANAGEMENT  
UNIVERSIDADE DE LISBOA

# ANÁLISE DE DADOS EM GRH

## *Aula 9: O modelo de regressão linear*

Por hoje é tudo. Até à próxima aula!