

Aula 7:

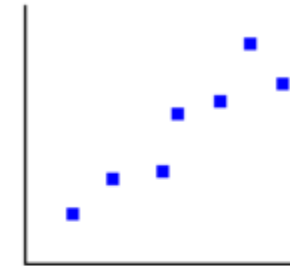
'Que fatores explicam a variação nos salários na organização?'

O modelo de regressão linear

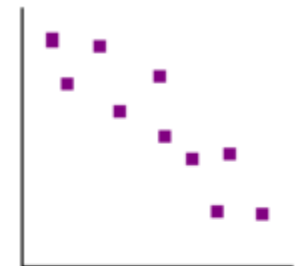
Docente: Daniela Craveiro
dcraveiro@iseg.ulisboa.pt

- **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**
 - **Estratégias de reporte de resultados dos testes de hipóteses**
 - **Parte Prática**
 - **Saber implementar o modelo de regressão linear**
 - **Saber interpretar os resultados do modelo de regressão linear**

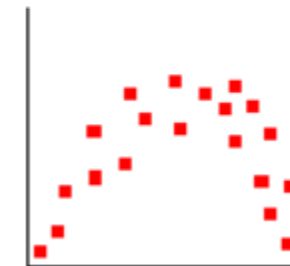
- **Coeficientes de Associação / Correlação**
 - Permitem aferir se existe uma relação sistemática entre duas variáveis
 - Permitem aferir a força da relação sistemática entre duas variáveis
 - Permitem identificar a direção da relação sistemática entre duas variáveis (correlação apenas)
- **Modelo de 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



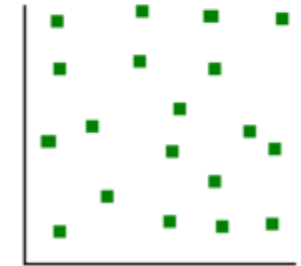
Positive linear relationship



Negative linear relationship



Non-linear relationship



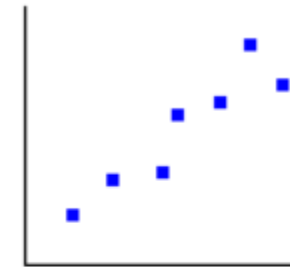
No relationship

Variável Independente:

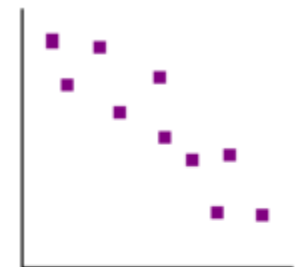
- A variável que assumimos/presumimos que influência/m outras variáveis
- Usada para explicar ou prever resultados (+/-) na variável dependente
- Define categorias ou valores que usamos para observar seu efeito sobre a variável dependente
- Pode ser considerada a causa ou o determinante
- Eixo x

Variável Dependente:

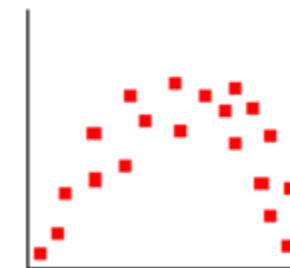
- A variável que estamos interessados em perceber como varia
- A variável que estamos interessados em perceber como varia em função da variável independente
- Pode ser considerada o resultado ou consequência do que está sendo estudado
- Eixo y



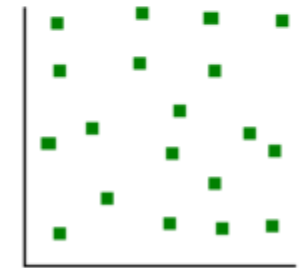
Positive linear relationship



Negative linear relationship



Non-linear relationship



No relationship

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

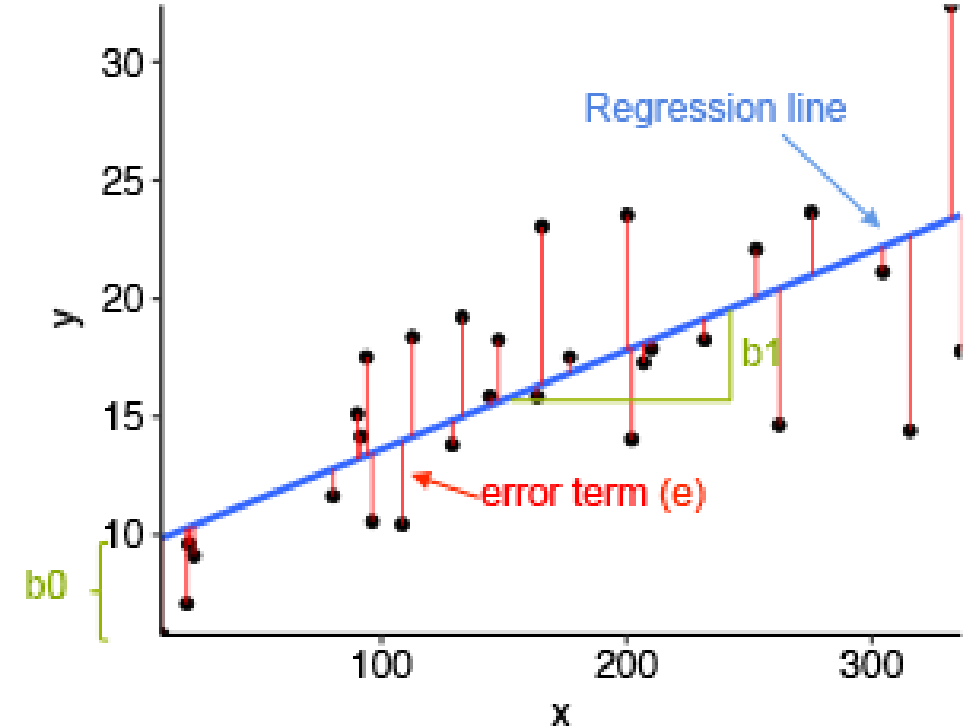
$$Y = \beta_0 + \beta_1 X + E$$

Y Variável Dependente

β_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)

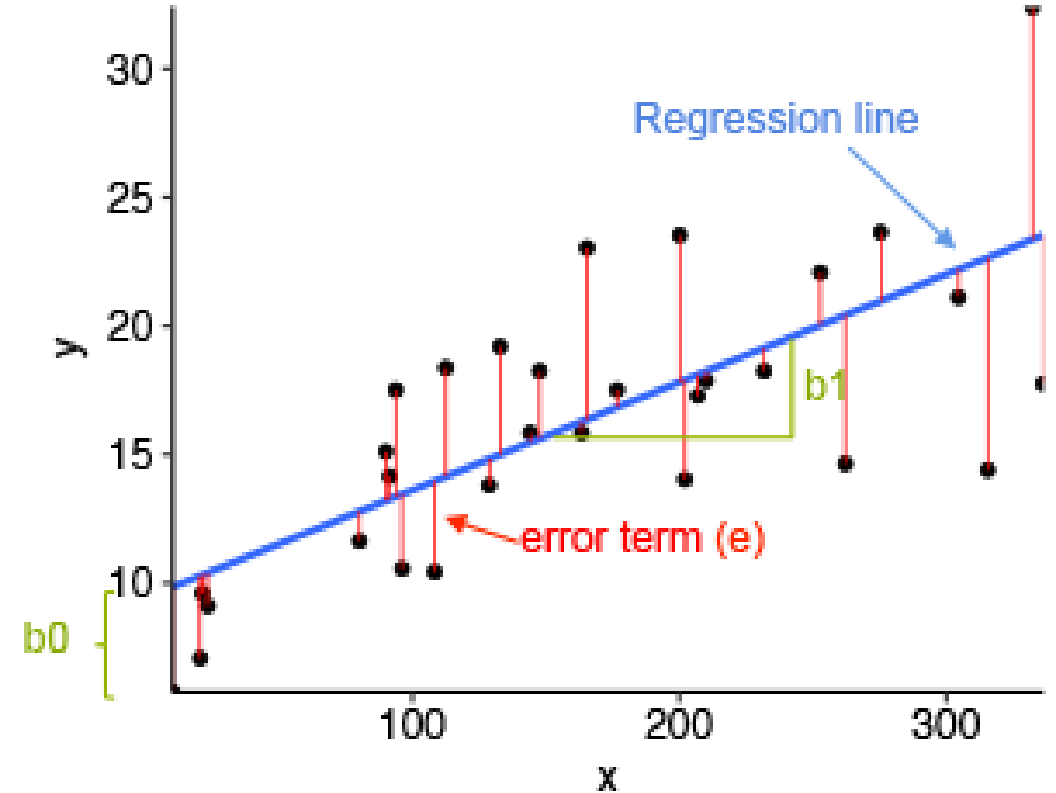
E Erro aleatório ou estocástico (reflete a influência de outros factores no no comportamento da variável Y que não podem ser explicadas linearmente pelo comportamento da variável X)



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

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

- A 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)**

=

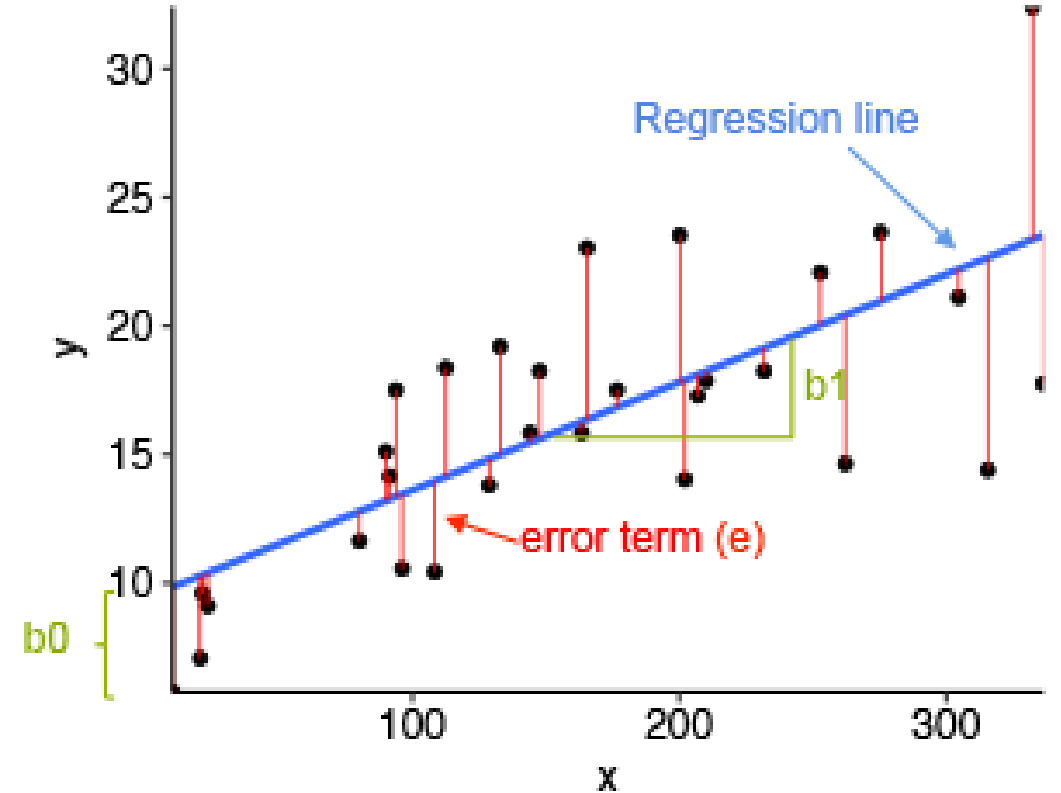
Varição explicada pelo modelo (SSR) = Soma dos Quadrados da Regressão

+

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

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

$$R^2 = SSR / SST$$



- 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^a

Modelo	Variáveis inseridas A	Variáveis removidas B	Método C
1	Age, Height ^b	.	Inserir

a. Variável Dependente: Weight

b. Todas as variáveis solicitadas inseridas.

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

- 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 β igual a 0.

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

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

ANOVA^a

Modelo		Soma dos Quadrados	df	Quadrado Médio	Z ^A	Sig. ^B
1	Regressão	692,823	2	346,411	15,953	,001 ^b
	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>

- 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 A	R quadrado ajustado B	Erro padrão da estimativa
1	,883 ^a	,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^a

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’

O que é que nos dizem estes coeficientes (^C)?

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^a

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	,261	,548		2,768	,022
Age			,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^a

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 ≤ 0.05

Grau de confiança = 99% → Sig. tem de ser ≤ 0.01

Grau de confiança = 90% → Sig. tem de ser ≤ 0.1

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

Coeficientes^a

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*

Modelo de Regressão Linear

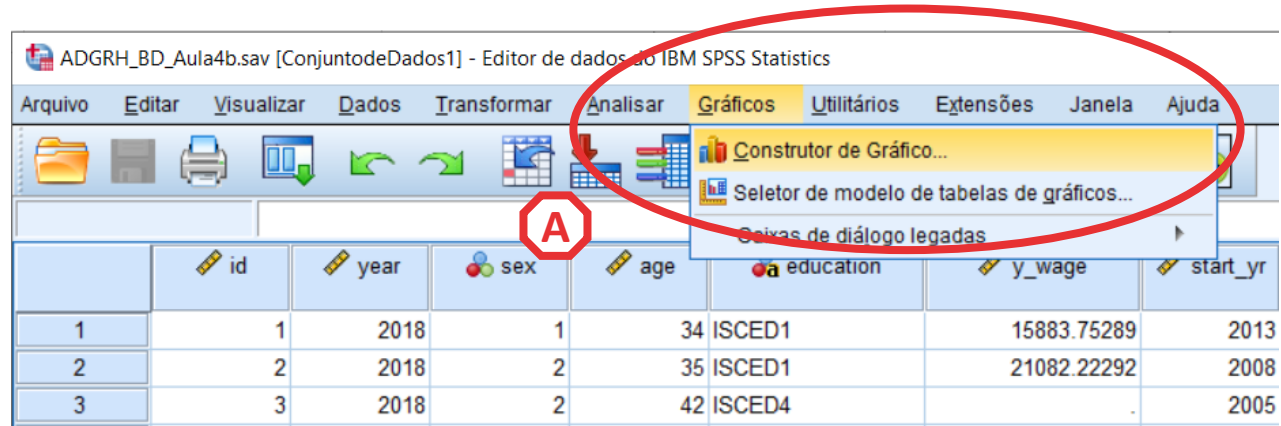
- **Objectivo:**
 - **Avaliar em que medida a distribuição dos salários na empresa variam em função da experiência dos trabalhadores e das avaliações que os seus supervisores**

Modelo de Regressão Linear

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

Histograma

- Seleccionar 'Gráficos' / 'Constructor de Gráfico'



Histograma

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

The screenshot shows the 'Construtor de gráfico' (Chart Builder) window. The title bar reads 'Construtor de gráfico'. Below the title bar, there are three tabs: 'Propriedades do elemento', 'Aparência do Gráfico', and 'Opções'. The main area is divided into several sections:

- Variáveis:** A list of variables including 'year', 'sex', 'age', 'education', 'start_yr', 'start_age', 'experience', 'y_wage', 'department', and 'hrs_week'. A red callout 'A' points to this list.
- Workspace:** A large empty area with instructions: 'Arraste um gráfico da Galeria aqui para usá-lo como seu ponto de início' and 'Clique na guia Elementos Básicos para construir um elemento de gráfico por elemento'. A red callout 'B' points to the 'Histograma' option in the gallery below.
- Galeria:** A section with tabs for 'Galeria', 'Elementos básicos', 'ID de elementos/ponto', and 'Títulos/Notas de rodapé'. Under the 'Galeria' tab, there is a list of chart types: 'Favoritos', 'Barra', 'Linha', 'Área', 'Setor/Polar', 'Dispersão/Ponto', 'Histograma', 'Auto Baixo', 'Boxplot', and 'Eixos duplos'. A red callout 'C' points to the 'Histograma' icon in the gallery.
- Buttons:** At the bottom, there are buttons for 'OK', 'Colar', 'Reconfigurar', 'Cancelar', and 'Ajuda'.

Histograma

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

Histograma Simples

Eixo Y?

Filtro?

Eixo X?

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:

Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

Título 1

Estatísticas

Variável:

Estatística: Histograma

Configurar parâmetros...

Exibir curva normal

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

Histograma

- Selecionar 'Gráficos' / 'Construtor de Gráfico' **A**
- Selecionar 'Histograma' **B**
- Duplo-clique no Histograma Simples **C**
- Selecionar a variável 'y_wage2' **E**
- E arrastar para o 'Eixo X' **E**
- Selecionar 'Exibir curva Normal' **F**
- Selecionar 'OK' **G**

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)

Histograma

Filtro?

y_wage2

Escolha entre:

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

OK Colar Reconfigurar Cancelar Ajuda

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de:

Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

Título 1

Estadísticas

Variável: y_wage2

Estatística: Histograma

Configurar parâmetros...

Exibir curva normal

Exibir barra de erros

Representação de Barras de

Intervalos de confiança

Nível (%): 95

Erro padrão

Multiplicador: 2

Desvio padrão

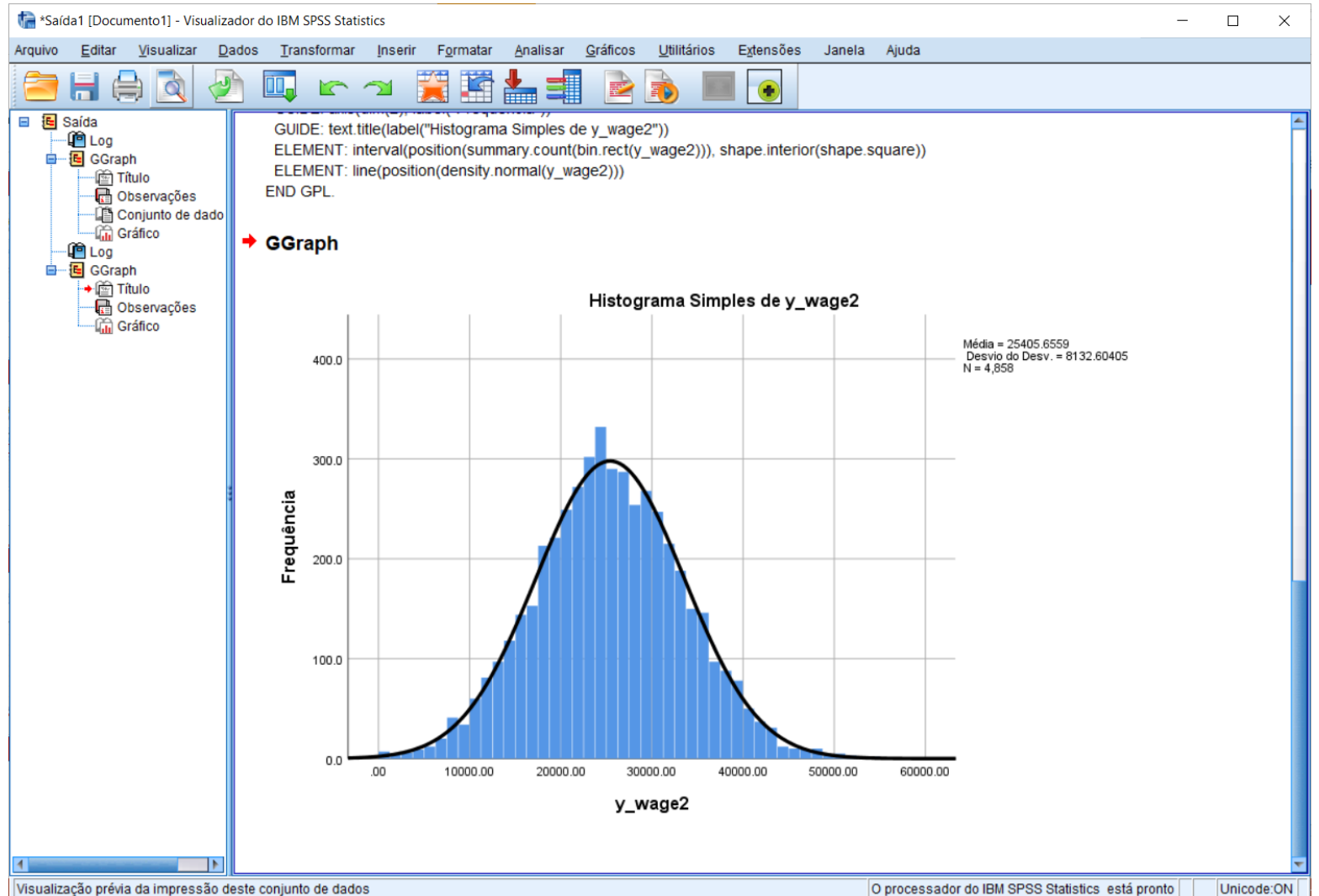
Multiplicador: 2

Estilo de barra:

Barra

Histograma

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



Modelo de Regressão Linear

- **Objectivo:**
 - **Detetar a existência de casos extremos**

Gráfico de Caixa

- Selecionar 'Gráficos' / 'Construtor de Gráfico'

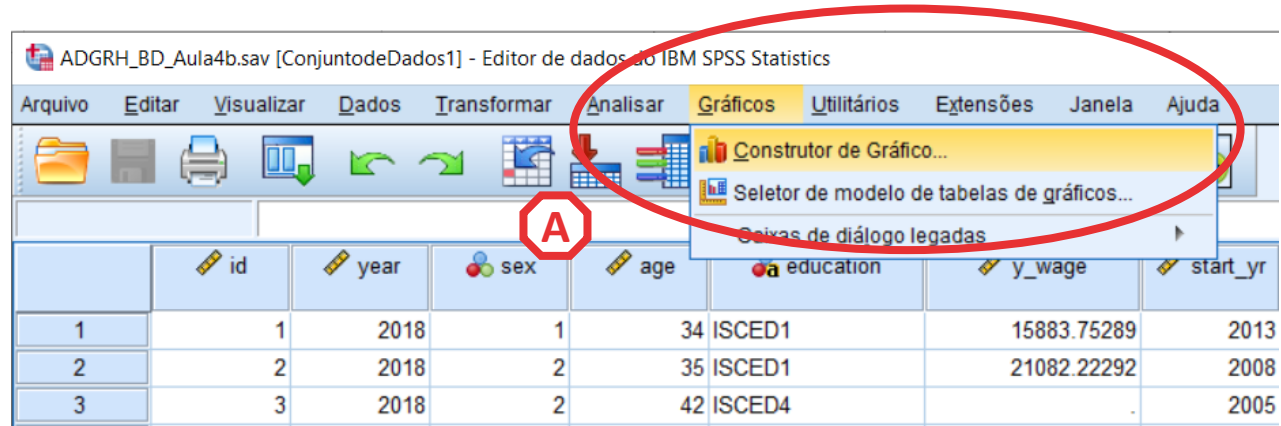


Gráfico de Caixa

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

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'

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'
- Selecionar 'OK'

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
- ...

Boxplot em 1-D de y_wage2

y_wage2

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...

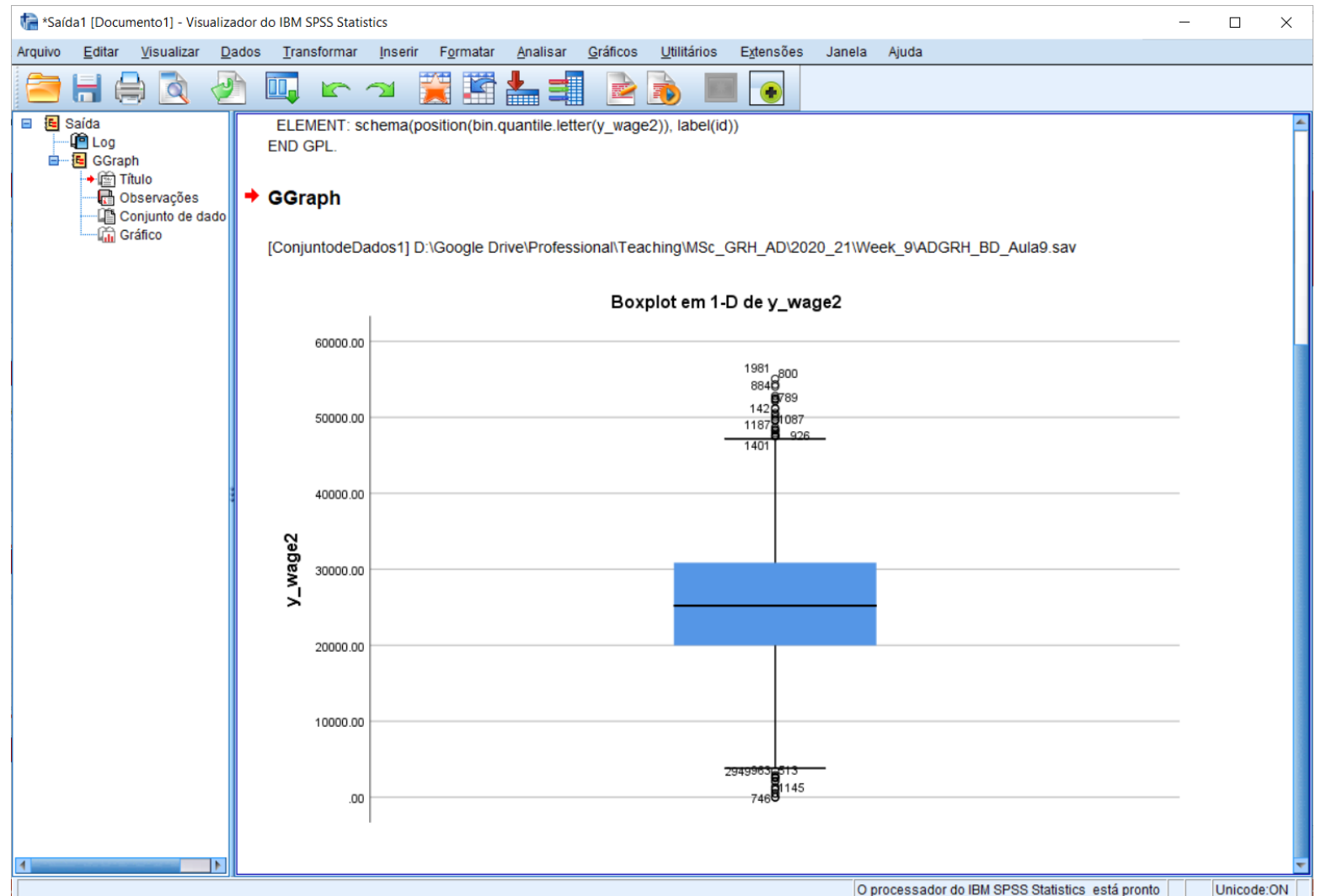
Exibir barra de erros

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'
- Sobre a forma como podemos lidar com estes casos extremos – ver a Aula 5!

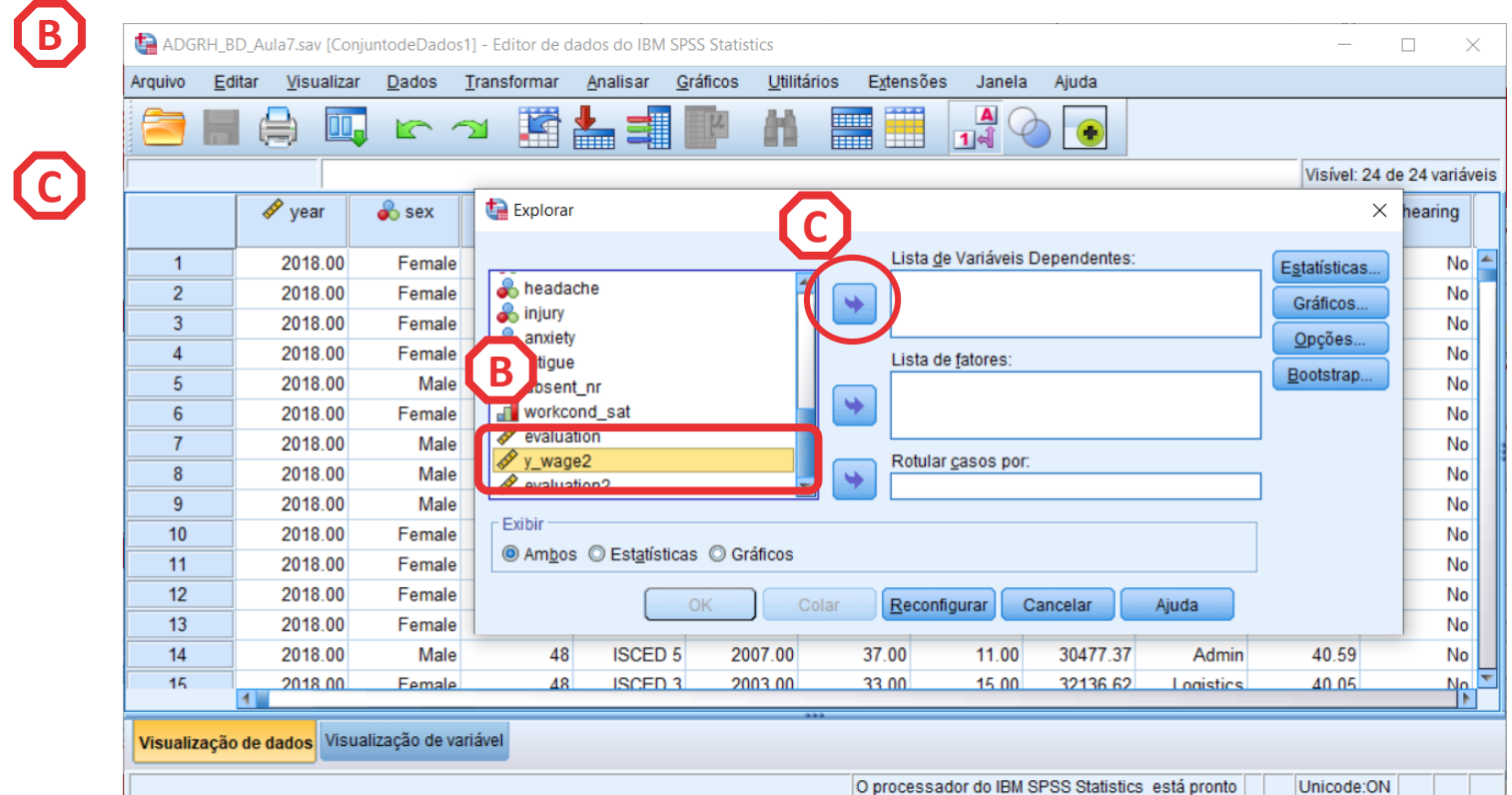
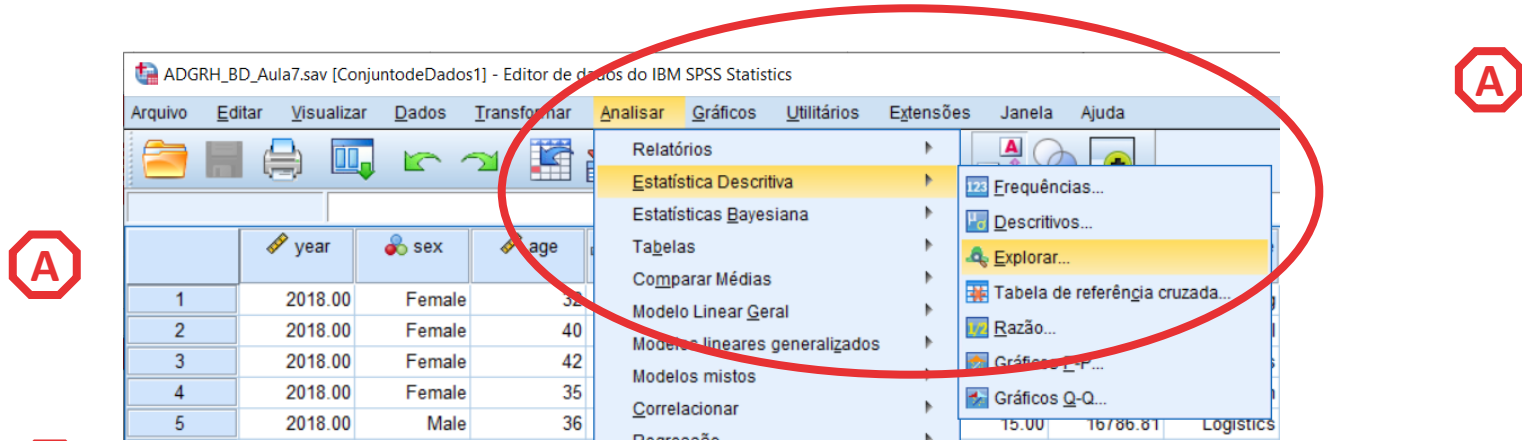


Modelo de Regressão Linear

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

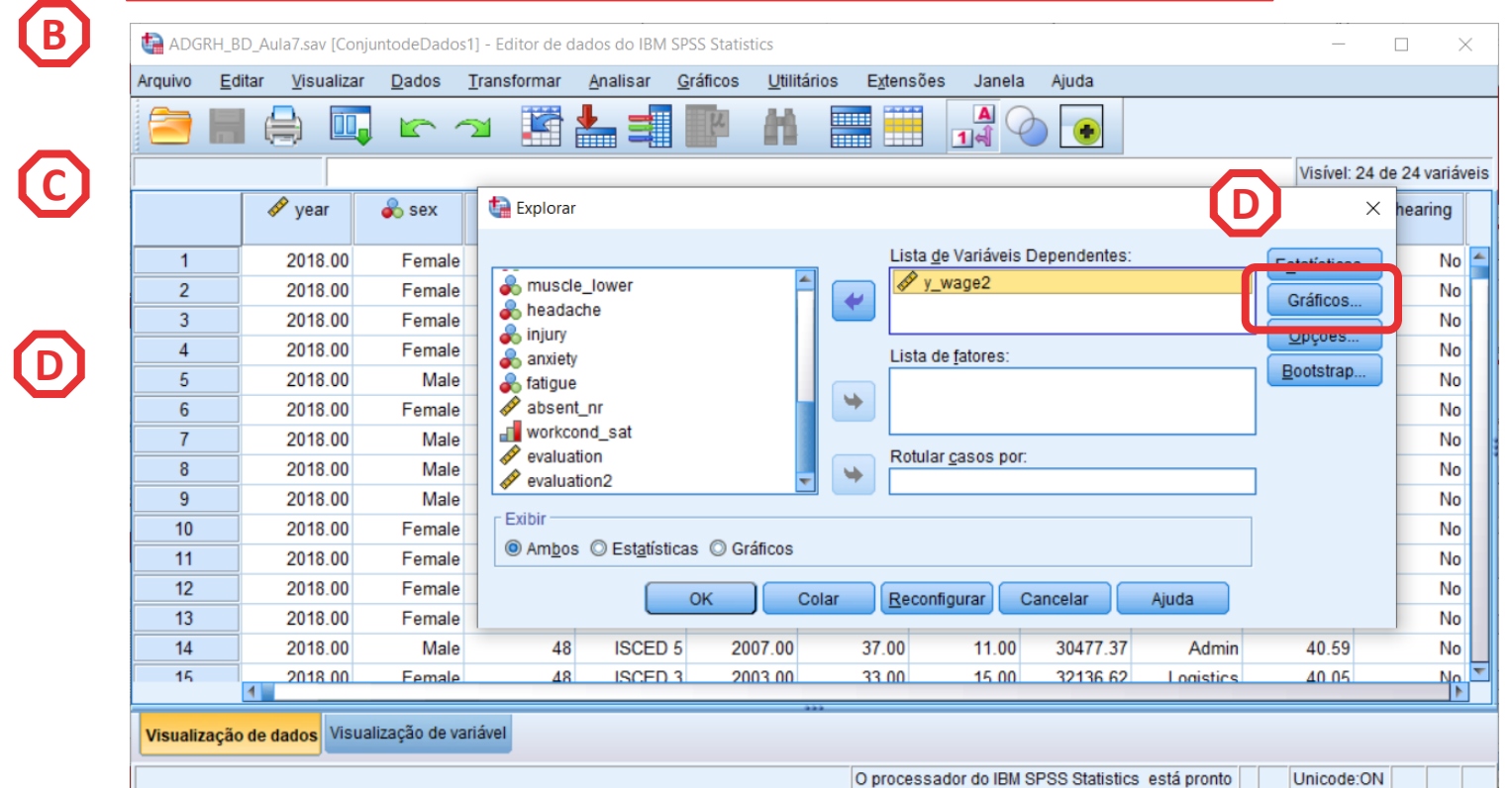
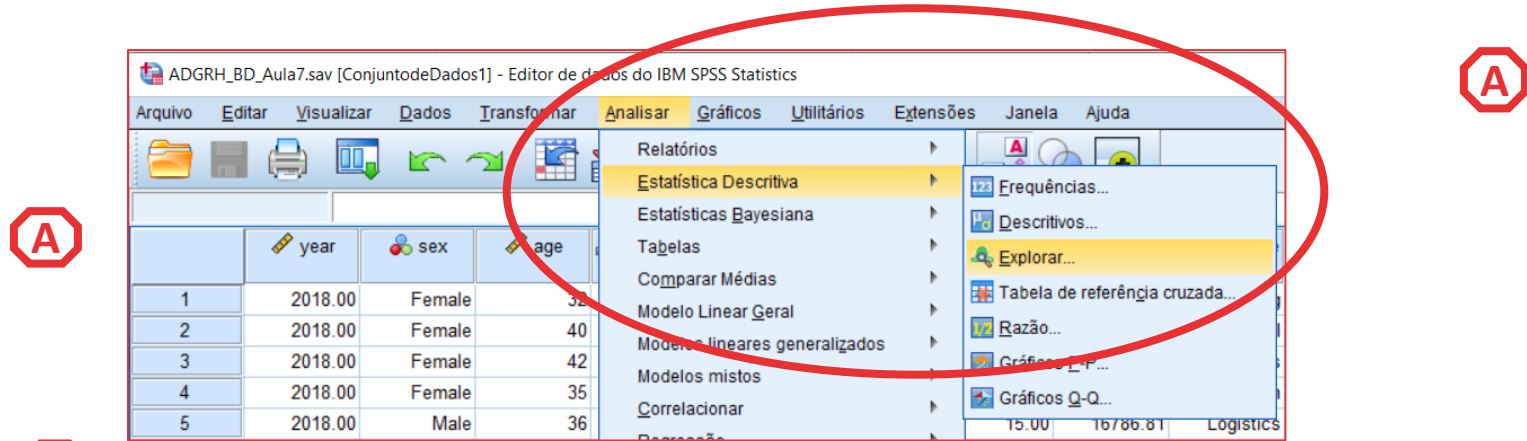
Teste de Shapiro-Wilk

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Explorar'
- Selecionar a variável 'y_wage2'
- Colocar na caixa 'Lista de Variáveis Dependentes'



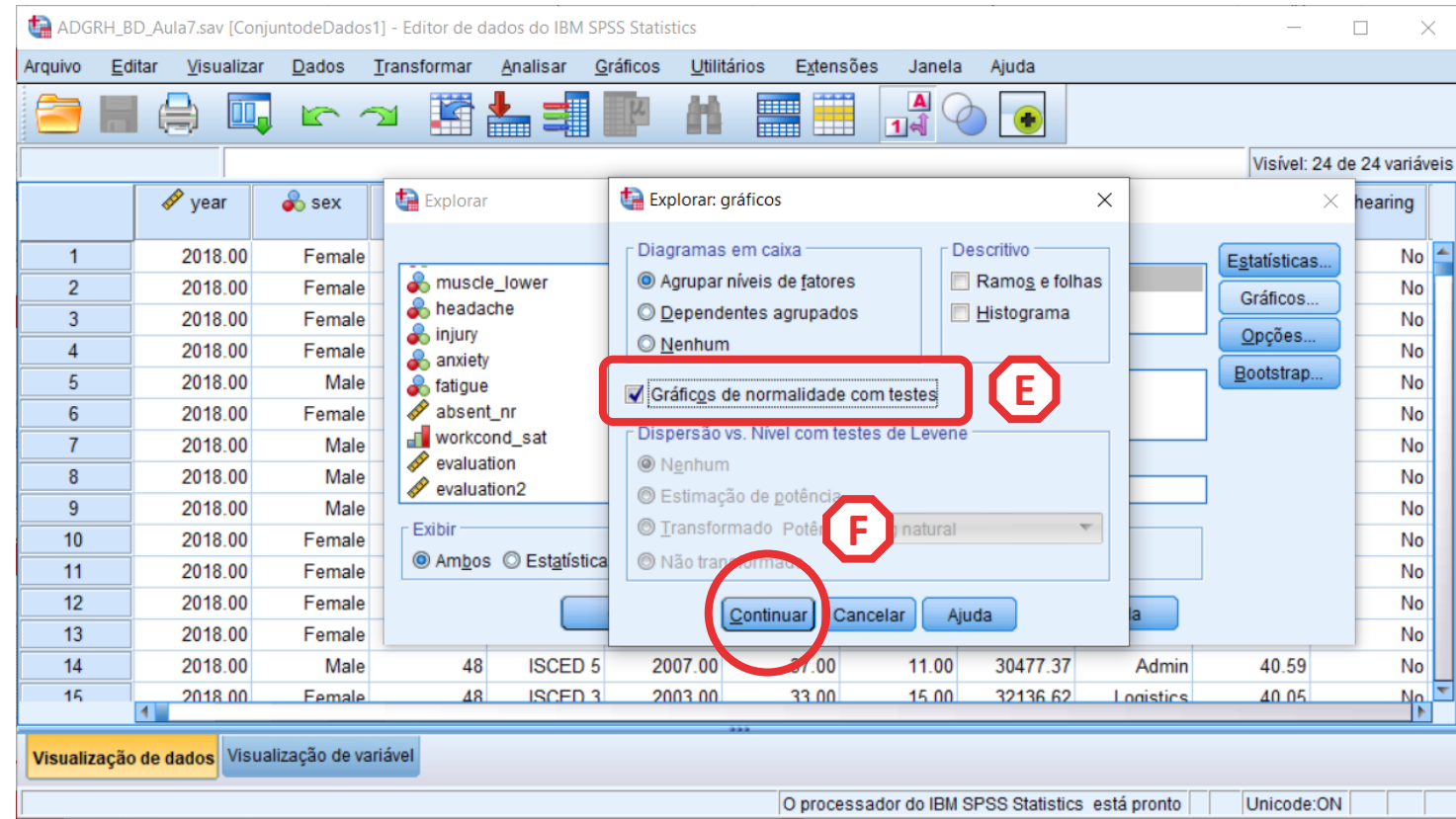
Teste de Shapiro-Wilk

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Explorar'
- Selecionar a variável 'y_wage2'
- Colocar na caixa 'Lista de Variáveis Dependentes'
- Selecionar 'Gráficos'



Teste de Shapiro-Wilk

- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Explorar' A
- Selecionar a variável 'y_wage2' B
- Colocar na caixa 'Lista de Variáveis Dependentes' C
- Selecionar 'Gráficos' D
- Selecionar "Gráficos de normalidade com testes" E
- Selecionar 'Continuar' 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



The screenshot shows the IBM SPSS Statistics interface with the 'Excluir Casos' dialog box open. The 'Excluir casos por método pairwise' option is selected under 'Valores omissos'. The 'Opções' button is highlighted with a red circle. The background shows a data table with columns for 'year' and 'sex'.

	year	sex									
1	2018.00	Female									
2	2018.00	Female									
3	2018.00	Female									
4	2018.00	Female									
5	2018.00	Male									
6	2018.00	Female									
7	2018.00	Male									
8	2018.00	Male									
9	2018.00	Male									
10	2018.00	Female									
11	2018.00	Female									
12	2018.00	Female									
13	2018.00	Female									
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

Teste de Shapiro-Wilk

- O resultado é publicado no 'Visualizador de Resultados'

INTERPRETAÇÃO:

- 'Sig'. ≤ 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

Saída

- Log
- Explorar
 - Título
 - Observações
 - Conjunto de dados
 - Resumo de processo
 - Descritivos
 - Testes de Normalidade
 - y_wage2
 - Título
 - Gráfico Q-Q n
 - Gráfico Q-Q n
 - Boxplot

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

y_wage2	Estatística	Erro	
		Estimado	Limite inferior
Média	25405.6559	116.68120	
95% Intervalo de Confiança para Média	25176.9080		
	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

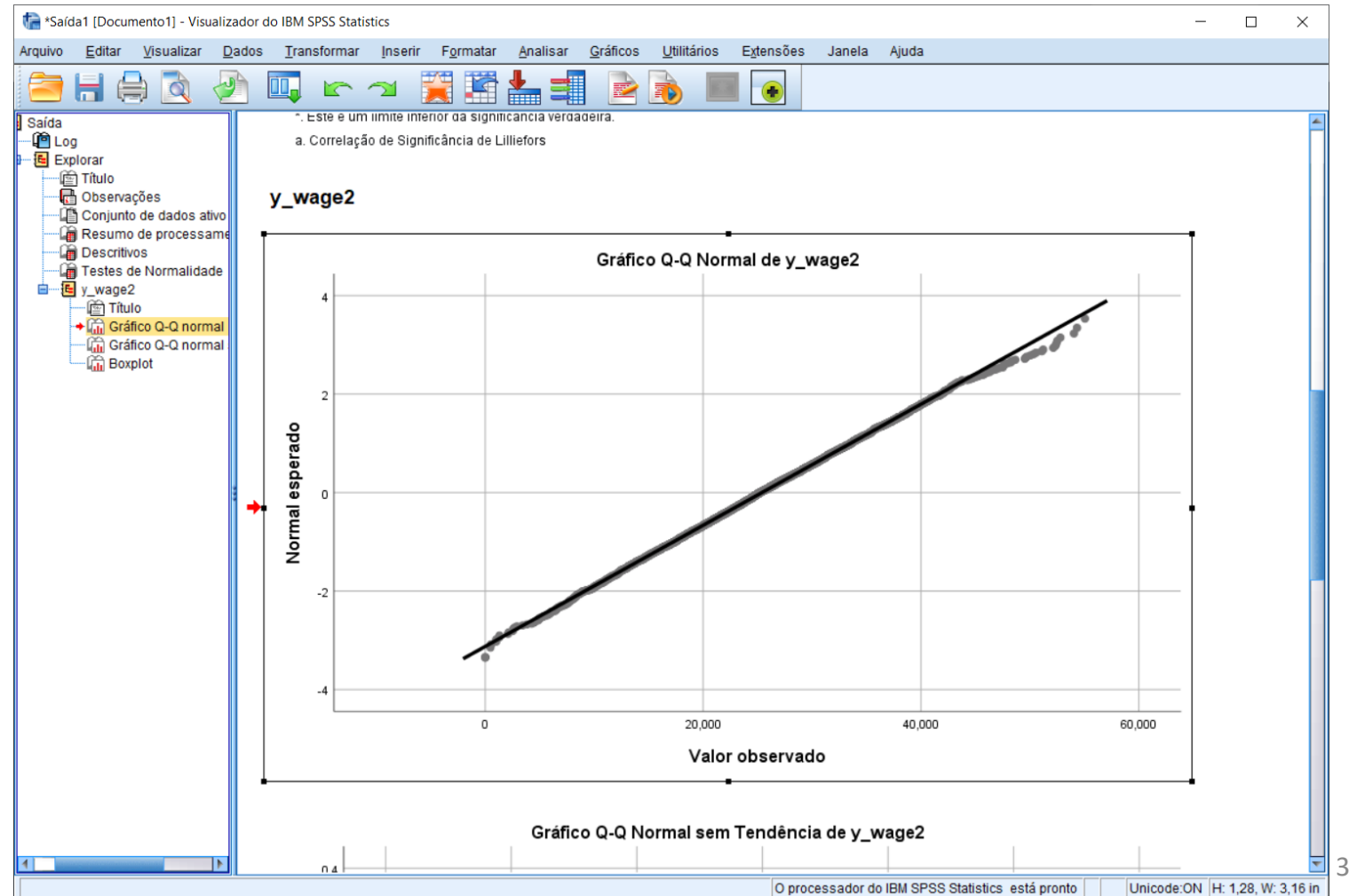
y_wage2	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estatística	df	Sig.	Estatística	df	Sig.
y_wage2	.011	4858	.200 [*]	.999	4858	.124

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

O processador do IBM SPSS Statistics está pronto Unicode:ON

Teste de Shapiro-Wilk

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



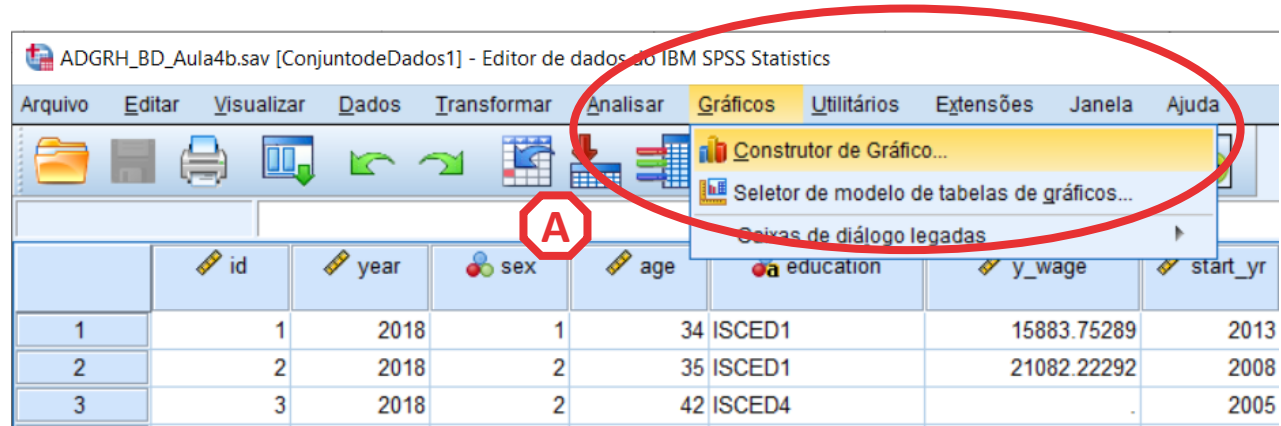
Implementação do modelo de regressão linear

3. Explorar as relações entre variáveis

Matriz de Dispersão

- Selecionar 'Gráficos' / 'Construtor de Gráfico'

A



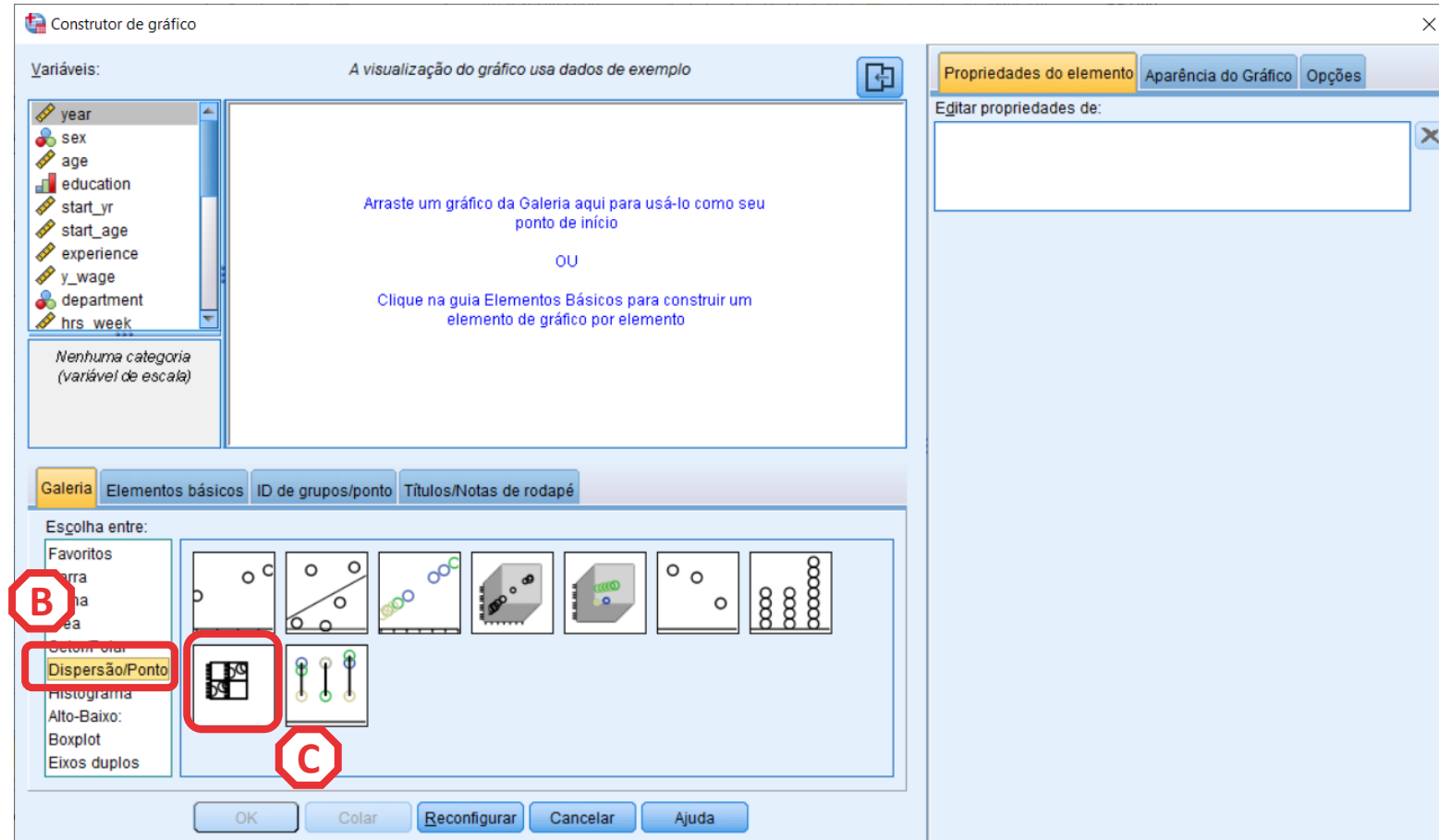
Matriz de Dispersão

- 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'

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

The screenshot displays the Minitab 'Construtor de gráfico' (Chart Builder) dialog box. The interface is divided into several sections:

- Variáveis:** A list of variables including 'muscle_lower', 'headache', 'injury', 'anxiety', 'fatigue', 'absent_nr', 'workcond_sat', 'evaluation', and 'y_wage2'. The variable 'y_wage2' is highlighted with a red box labeled 'D'.
- Matriz de Gráfico Disperso:** A central workspace showing a 2x2 grid of scatter plots. A red box labeled 'E' highlights the 'Matriz de dispersão?' box at the bottom of this workspace.
- Galeria:** A gallery of chart types including 'Barra', 'Linha', 'Área', 'Setor/Polar', 'Dispersão/Ponto', 'Histograma', 'Alto-Baixo', 'Boxplot', and 'Eixos duplos'. The 'Dispersão/Ponto' option is selected.
- Propriedades do elemento:** A panel on the right showing the properties for the selected chart type. It includes options for 'Exibir barra de erros' and 'Representação de Barras de Erros'.

Red octagonal callouts A through E indicate the steps described in the text:

- A:** Points to the 'Construtor de gráfico' title bar.
- B:** Points to the 'y_wage2' variable in the 'Variáveis' list.
- C:** Points to the 'Matriz de Gráfico Disperso' chart type in the gallery.
- D:** Points to the 'y_wage2' variable being dragged.
- E:** Points to the 'Matriz de dispersão?' box in the workspace.

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

Multiplicador: 2

Desvio padrão

Multiplicador: 2

Variáveis de matriz

Ordem:

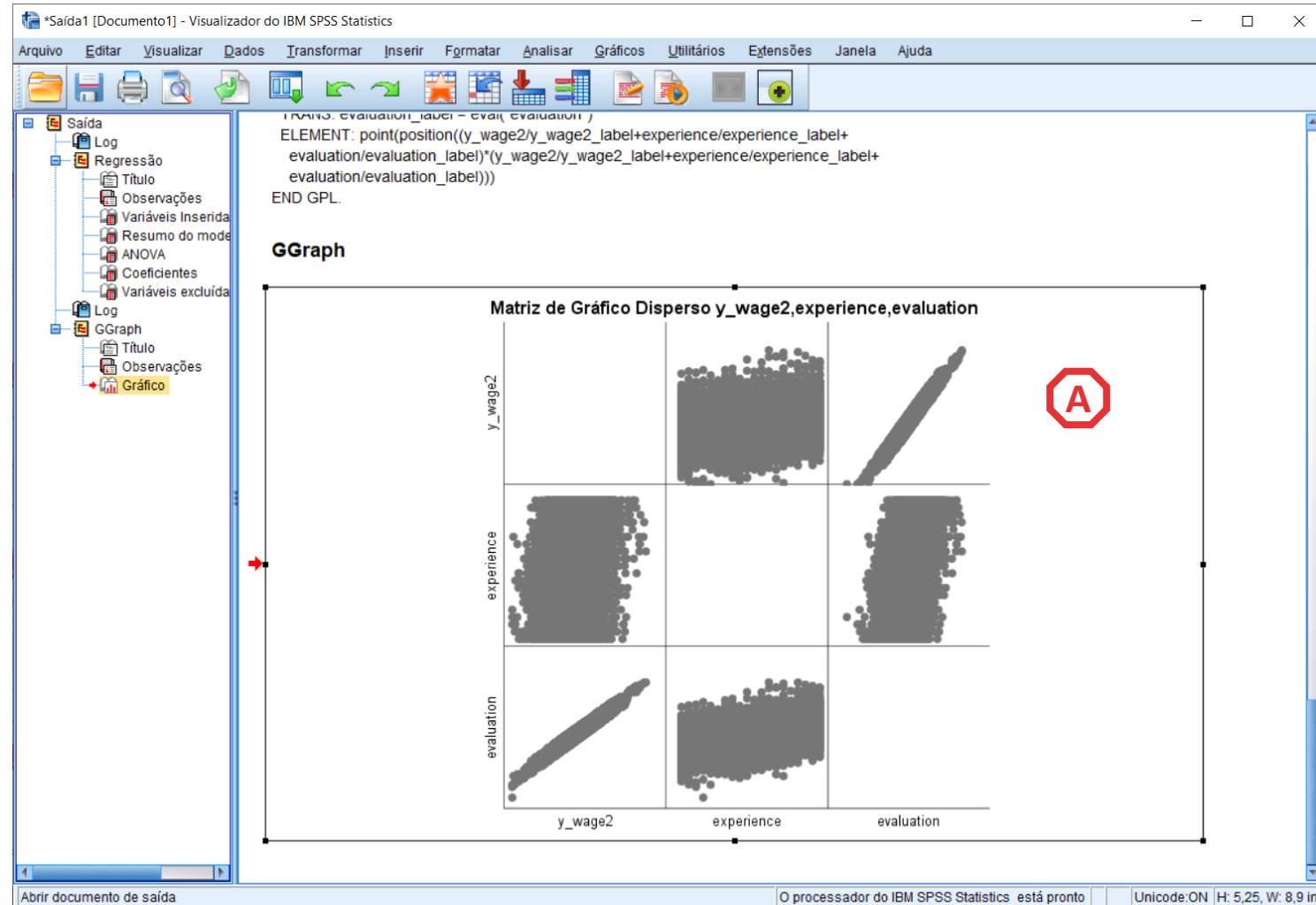
- y_wage2
- experience
- evaluation

Totál Subgrupos

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

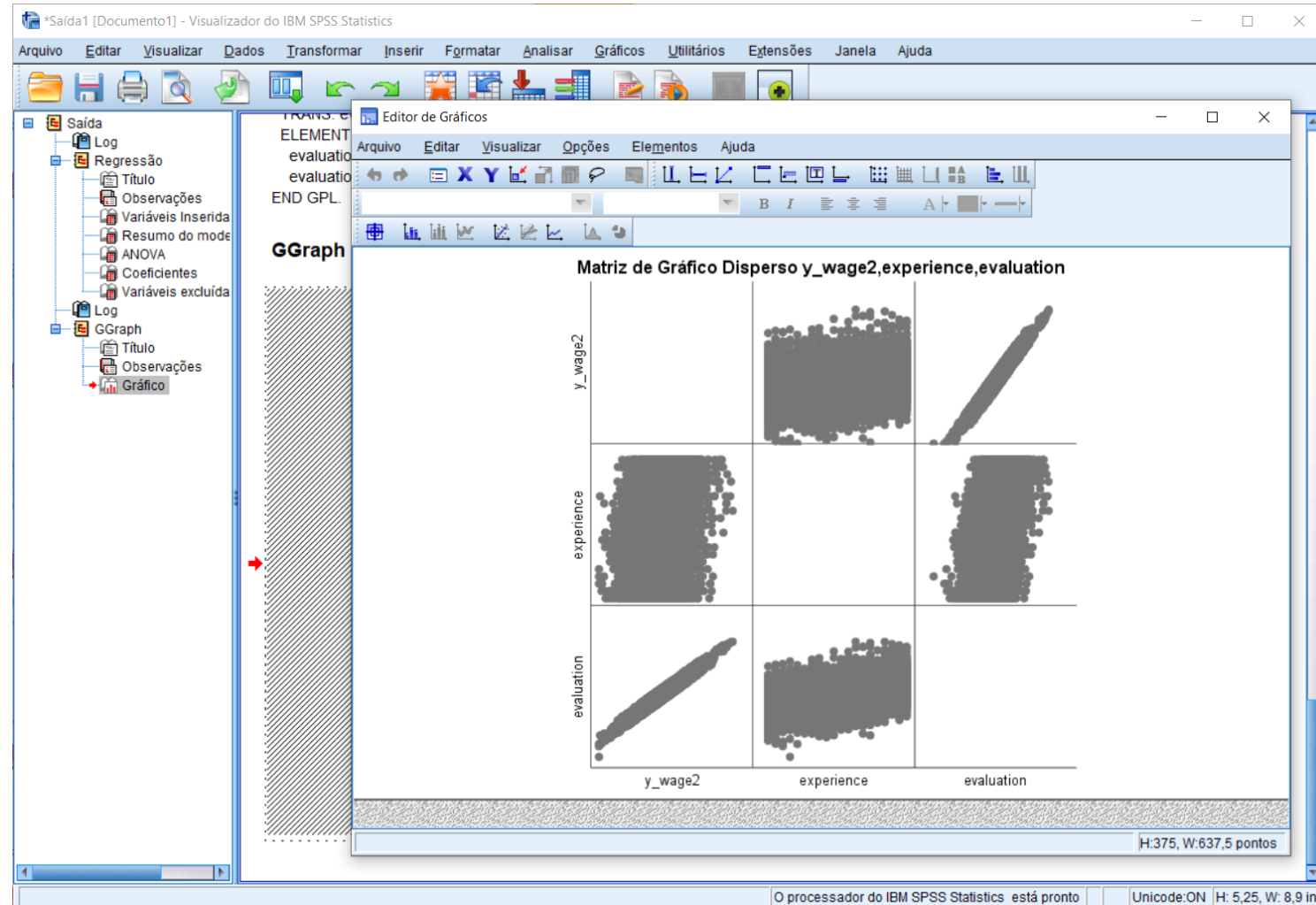
Matriz de Dispersão

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



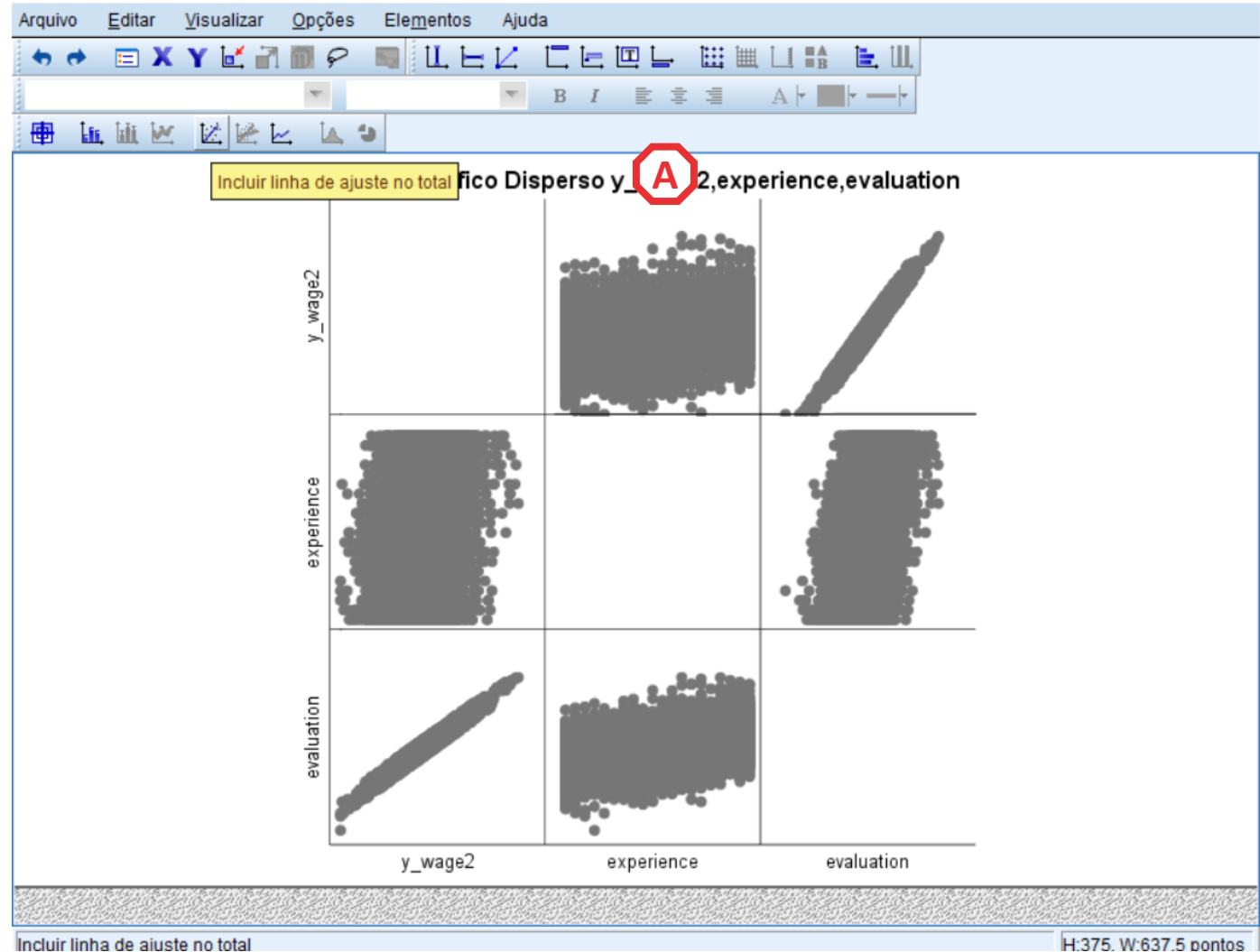
Matriz de Dispersão

- 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'



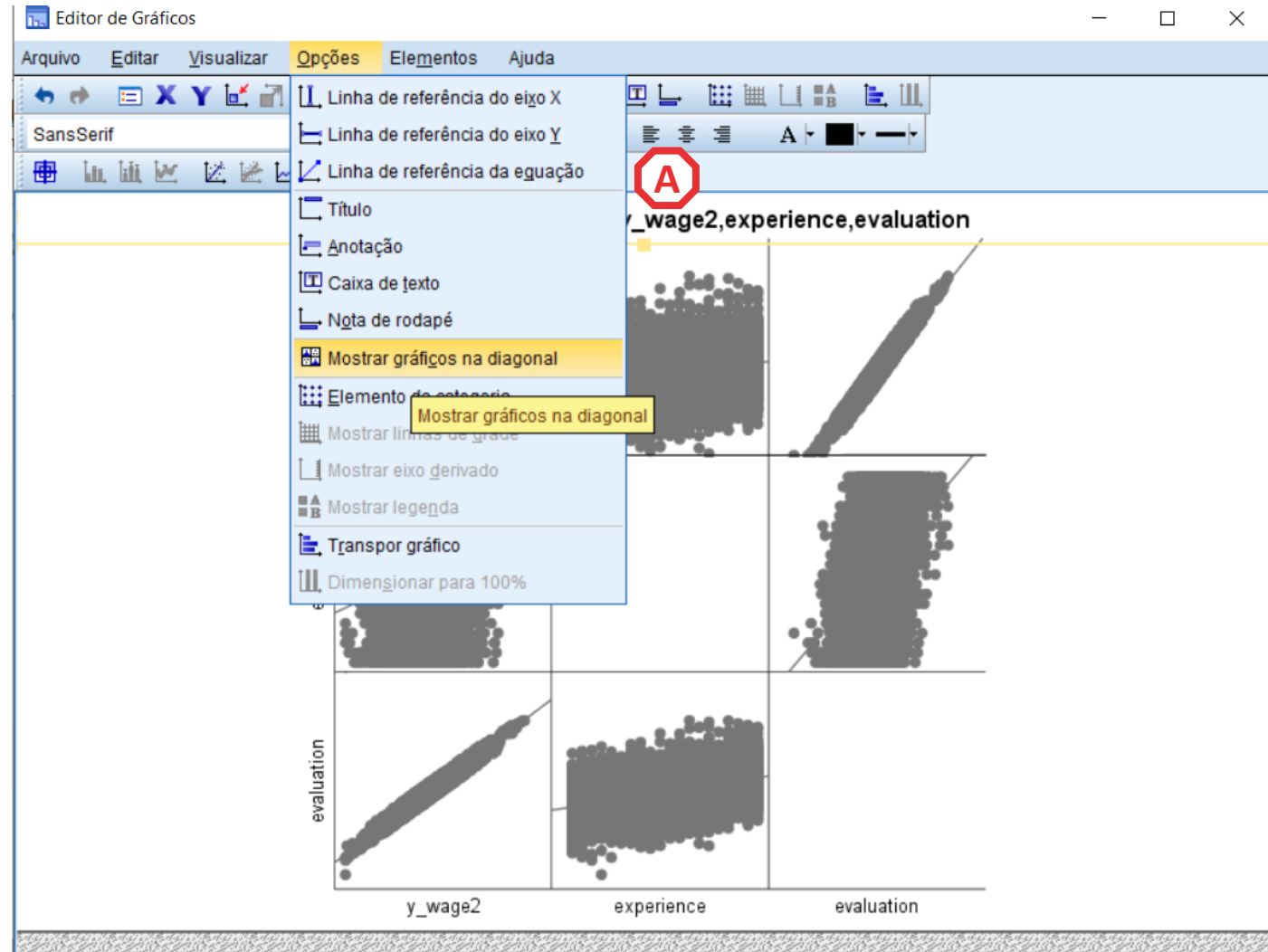
Matriz de Dispersão

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



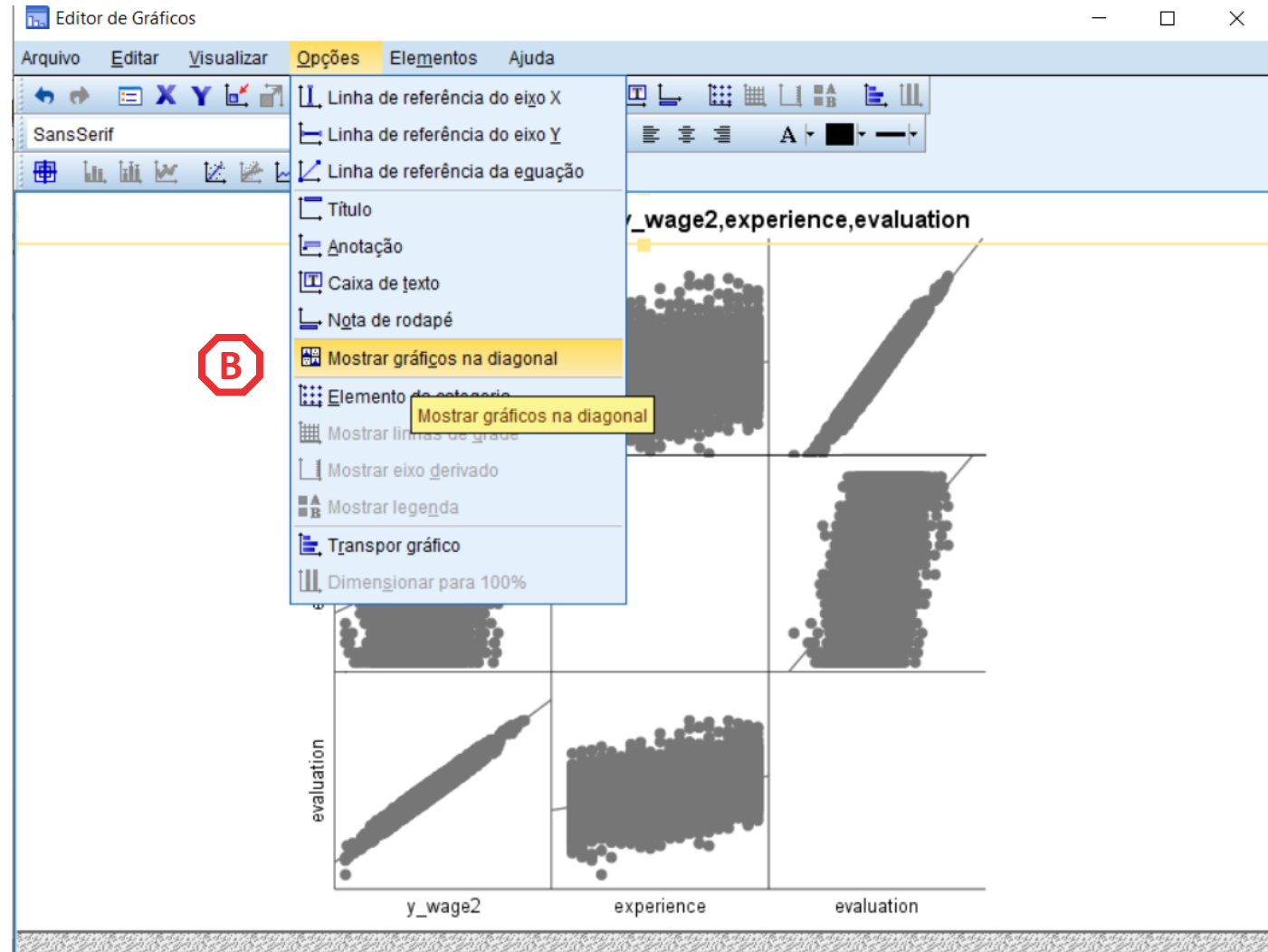
Matriz de Dispersão

- 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'



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

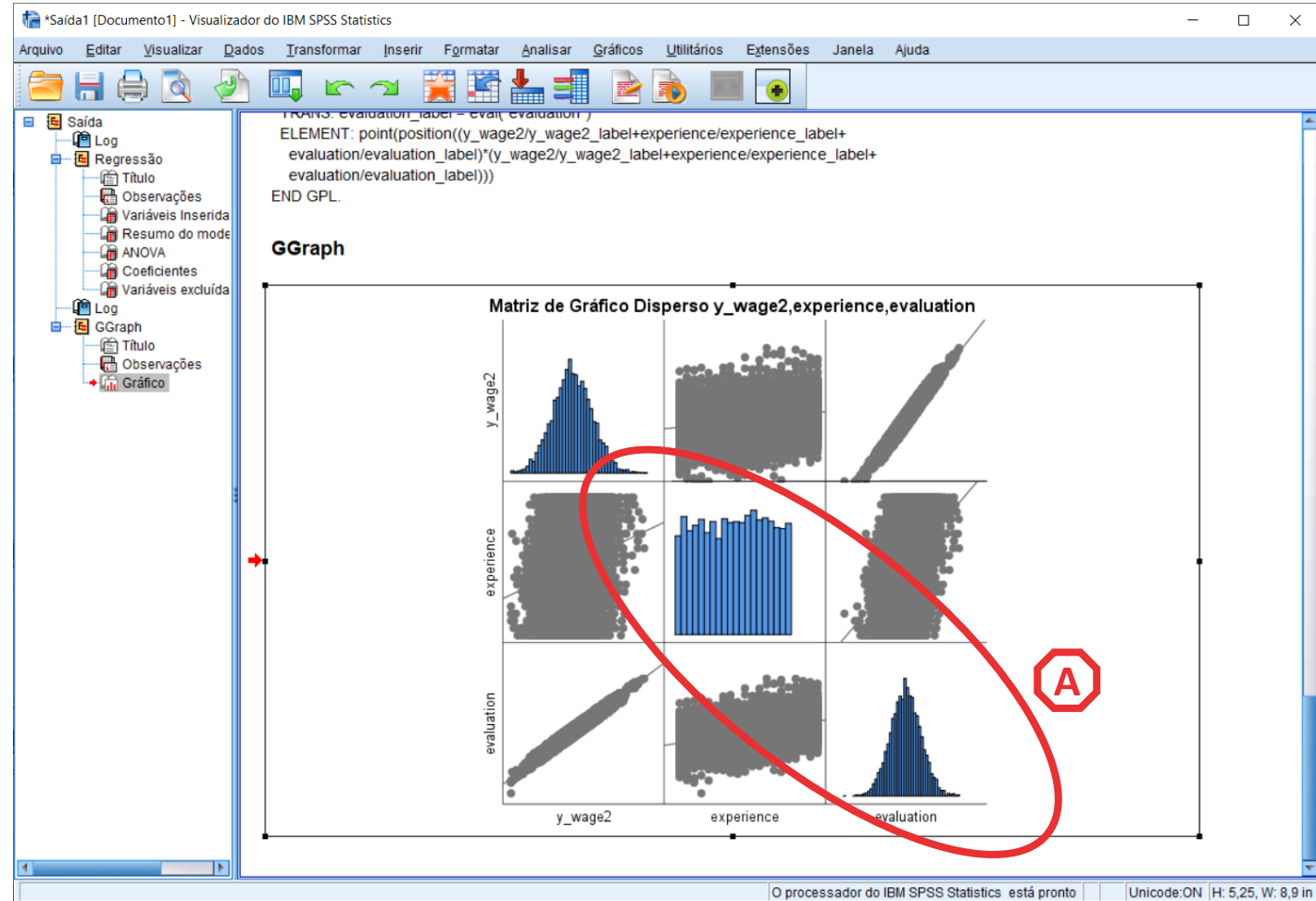
C

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 the 'Preenchimento e borda' tab. The dialog has a 'Visualização prévia' section, a 'Cor' section with 'Preenchimento' (85, 150, 230) and 'Borda' (0, 0, 0) options, and an 'Estilo da borda' section with 'Ponderação' (1), 'Estilo', and 'Extremidades' (Agrupado) options. The 'Fechar' button is circled in red. The status bar at the bottom right shows 'H:375, W:637,5 pontos'.

C

Matriz de Dispersão

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

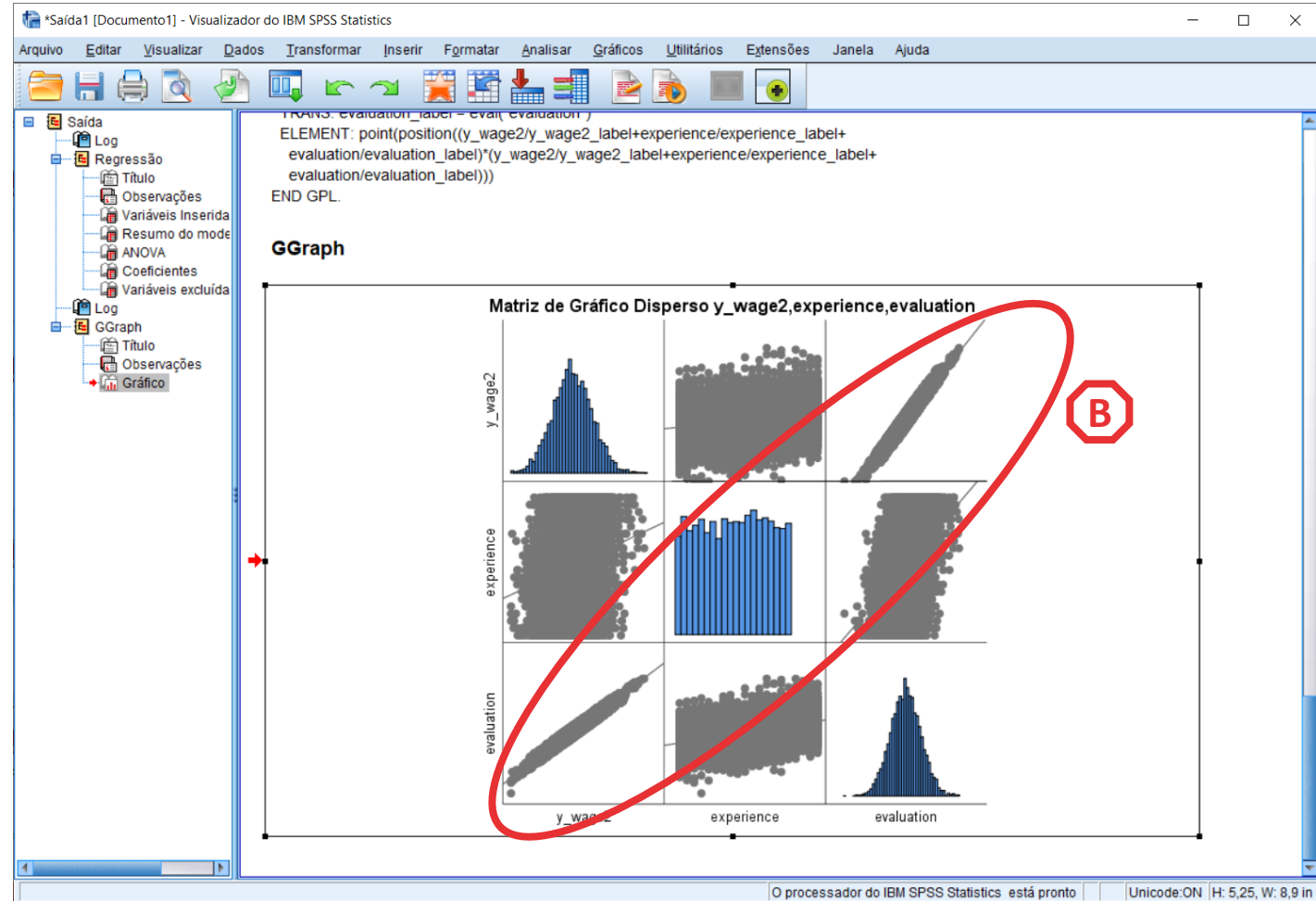


Matriz de Dispersão

- 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



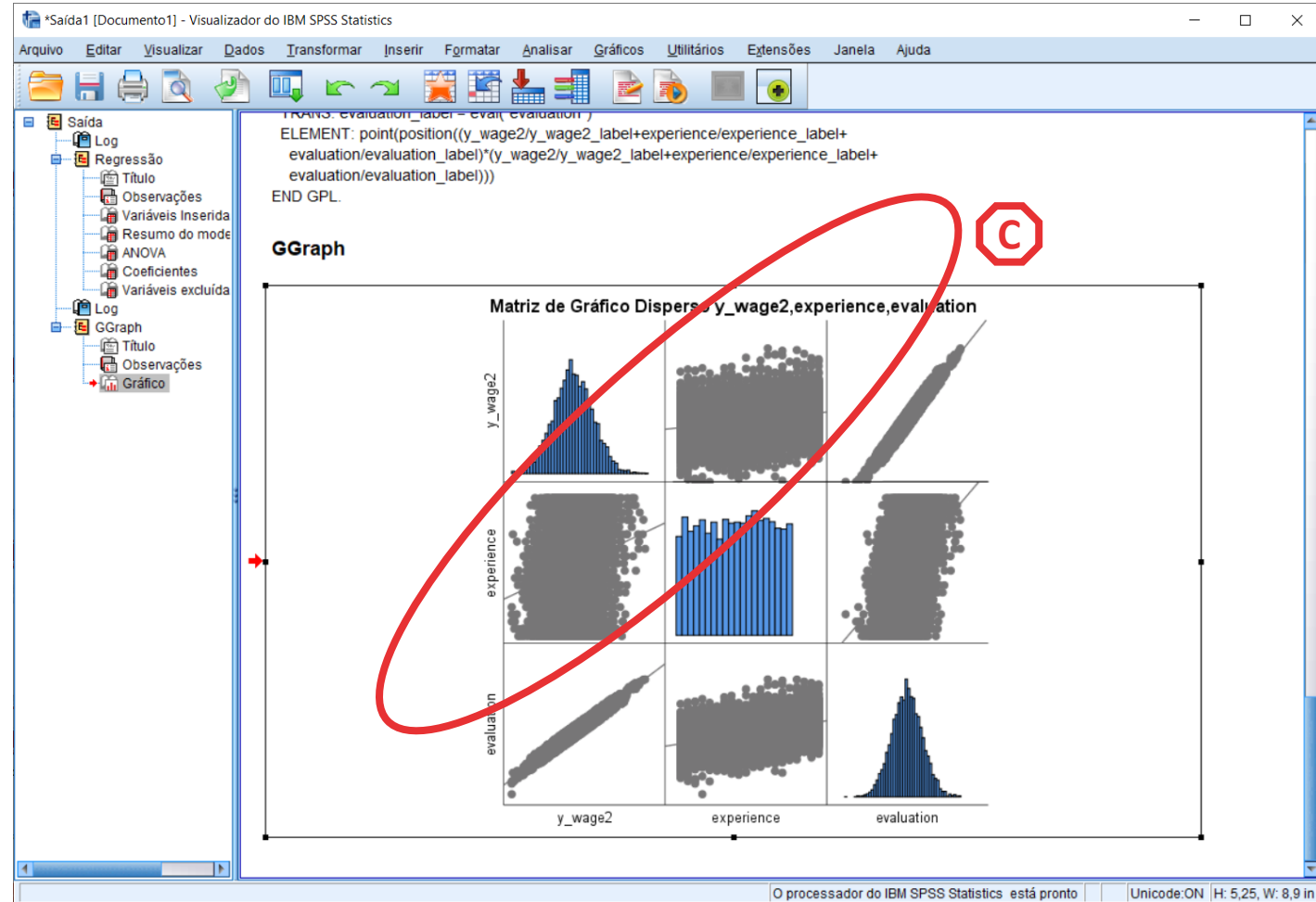
Matriz de Dispersão

- 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



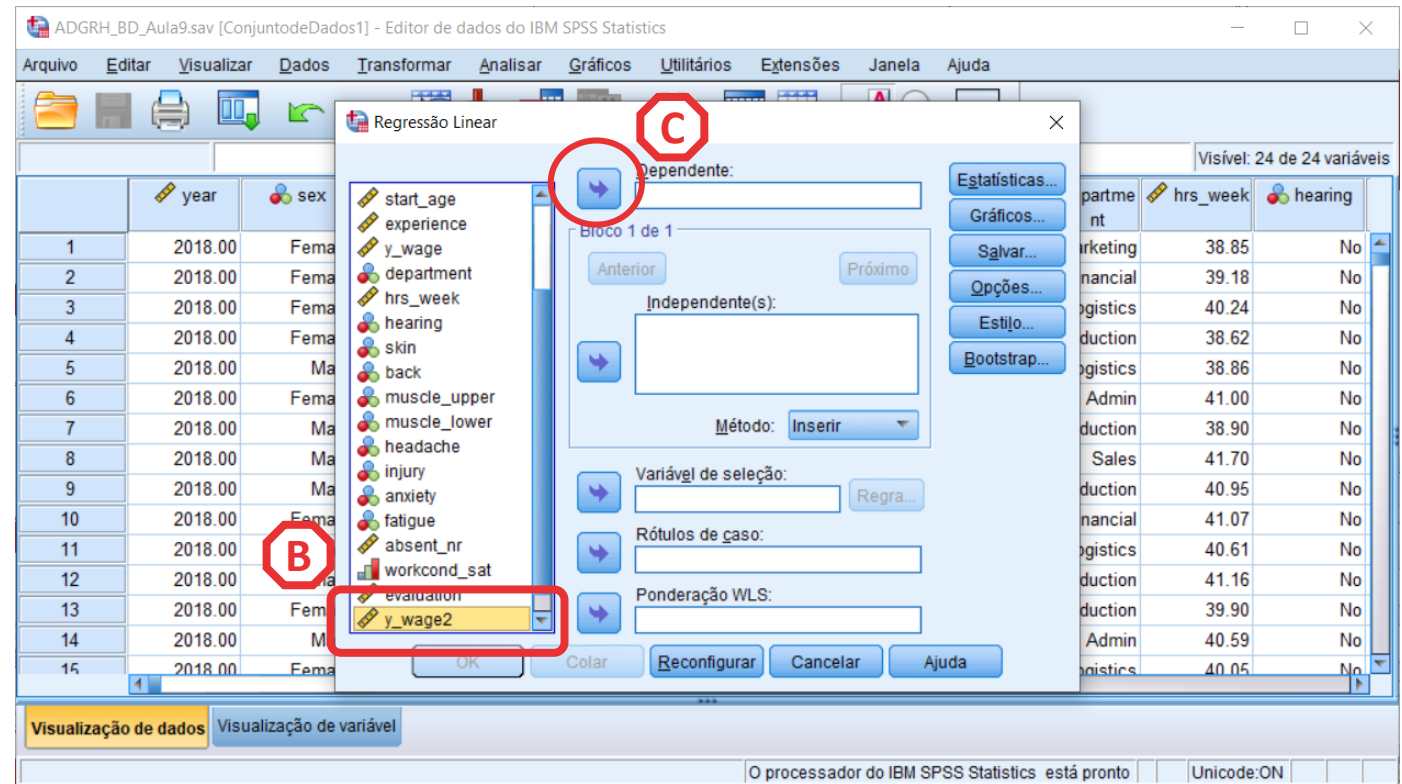
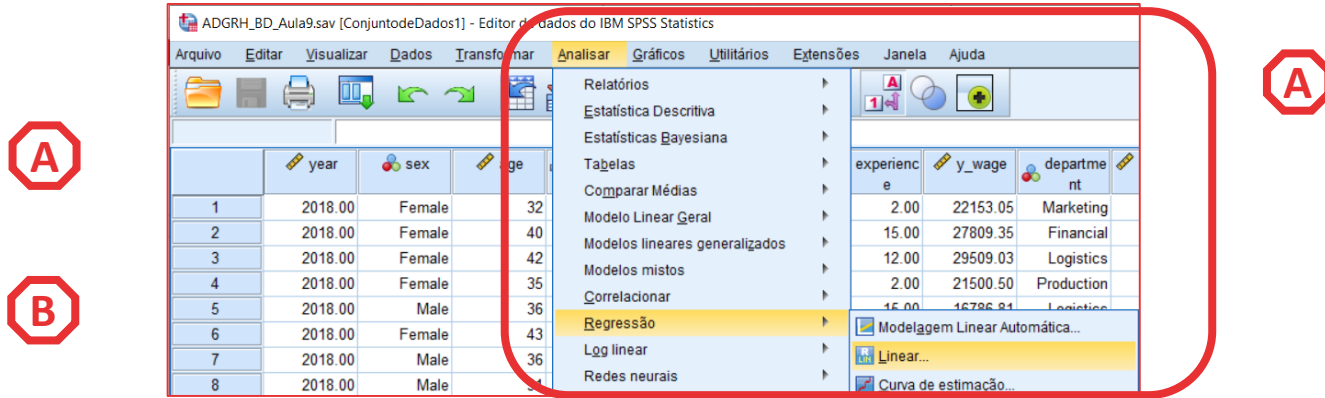
Implementação do modelo de regressão linear

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

Regressão Linear

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y_wage'
- Colocar na caixa 'Dependente'

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



Regressão Linear

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y_wage'
- Colocar na caixa 'Dependente'

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

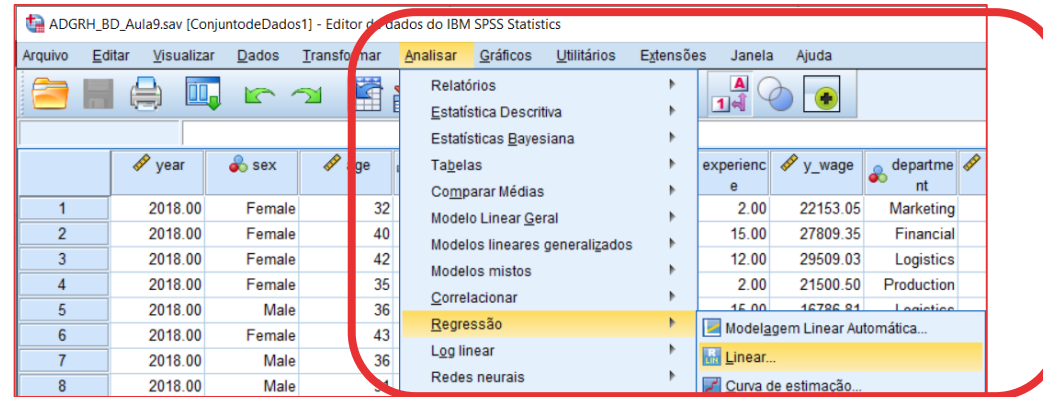
- Selecionar botão 'Estatísticas'

A

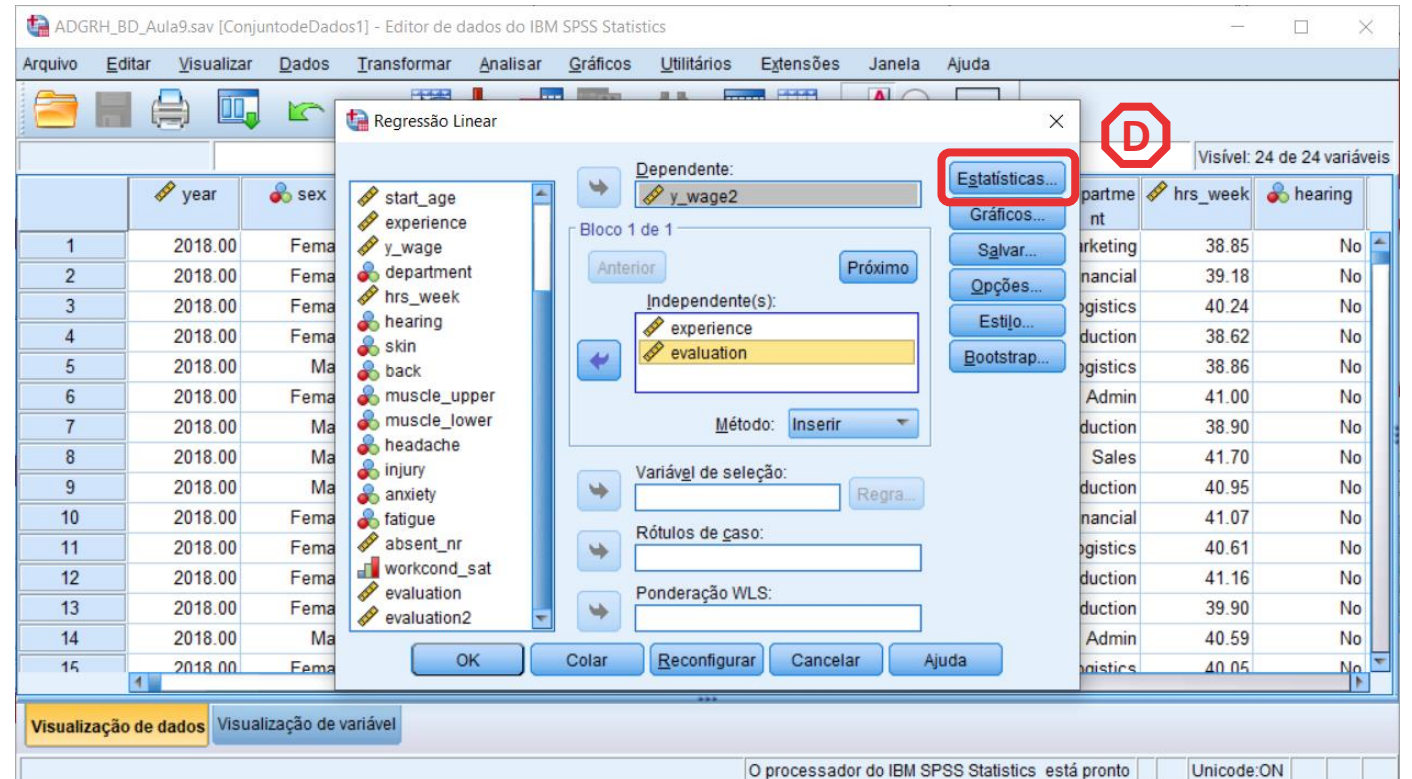
B

C

D



A



D

Regressão Linear

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

- Selecionar a variável 'y_wage'

- Colocar na caixa 'Dependente'

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

- Selecionar botão 'Estatísticas'

- Selecionar 'Estimativas'

- Selecionar 'Ajuste do modelo'

- Selecionar 'Continuar'/'OK'

A

B

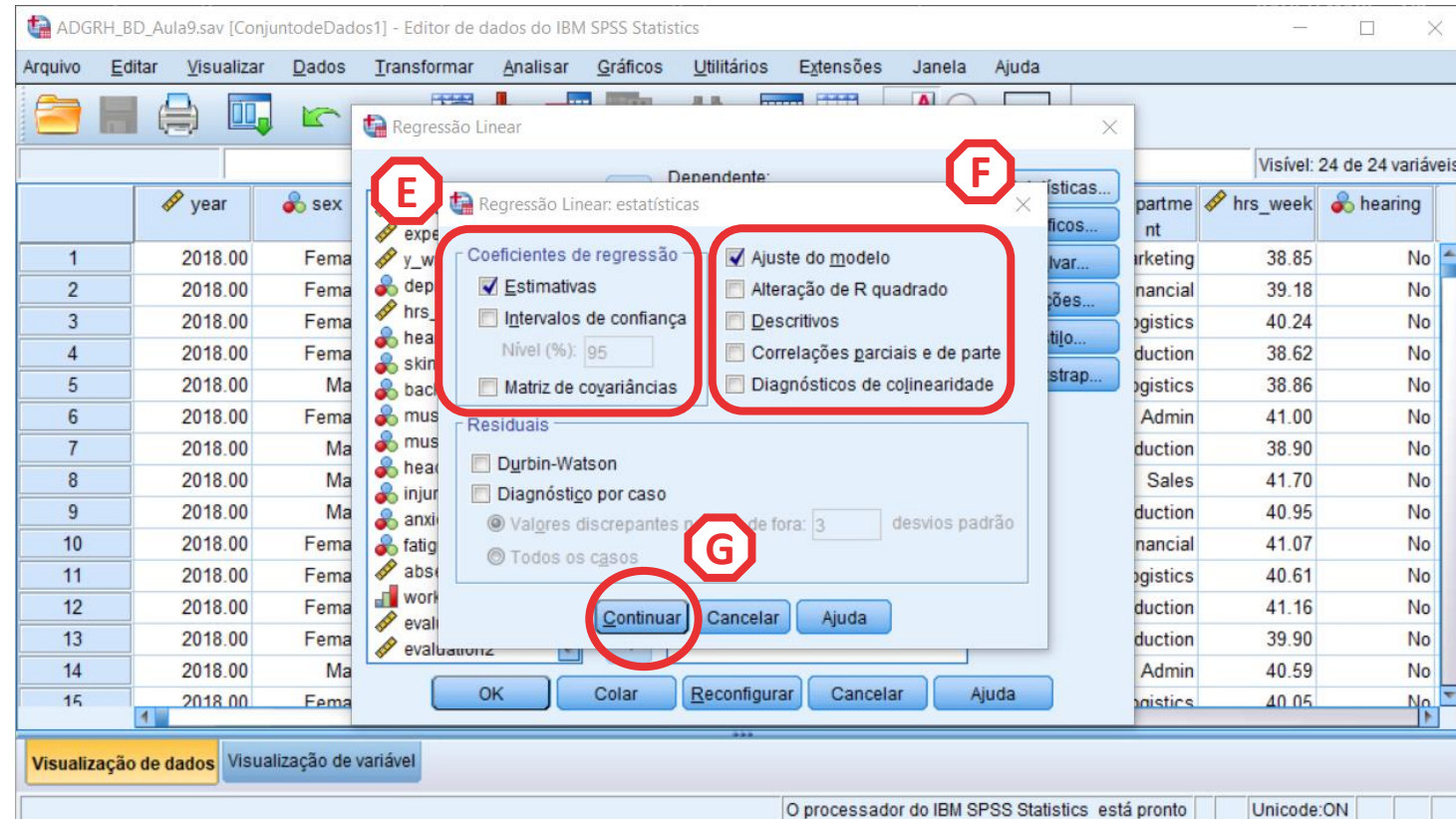
C

D

E

F

G



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^a

Modelo	Variáveis inseridas	Variáveis removidas	Método
1	evaluation, experience ^b		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 ^a	.997	.997	473.19356

a. Preditores: (Constante), evaluation, experience

ANOVA^a

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

a. Variável Dependente: y_wage2
b. Preditores: (Constante), evaluation, experience

Coefficientes^a

Modelo	Coefficients não padronizados		Coefficients 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_wage'

- Quais são as variáveis independentes?

experiência

avaliação dos trabalhadores

- Alguma variável foi excluída?

Não

Variáveis Inseridas/Removidas^a

Modelo	Variáveis inseridas	Variáveis removidas	Método
1	evaluation, experience ^b	.	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 ^a				
Modelo		Soma dos Quadrados	df	Quadrado Médio	Z	Sig.
1	Regressão	3,202E+11	2	1,601E+11	714903,694	,000 ^b
	Resíduo	1087093458	4855	223912,144		
	Total	3,212E+11	4857			

a. Variável Dependente: y_wage2

b. Preditores: (Constante), experience, evaluation

Regressão Linear

- Qual é o poder explicativo do nosso?

O R^2 é de .99

*→ O modelo explica 70% da
variação dos salários na
organização*

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



Resumo do modelo

Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa
1	,998 ^a	,997	,997	473,19356

a. Preditores: (Constante), experience, evaluation

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%
- Experience com efeito negativo
- Avaliação com efeito positivo
- O efeito da variável 'evaluation' sobre o salário anual dos trabalhadores é superior ao efeito da variável 'experience'.

A

B

Coefficientes^a

Modelo		Coefficients não padronizados		Coefficients 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 3354 Euros.



Mas não estamos a controlar o efeito de outras variáveis relevantes...

Coefficientes^a

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

Implementação do modelo de regressão linear

4. Adicionar IVs nominais

Adicionar IVs nominais

- **Objetivo:**
 - **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)

Basicamente transformar uma variável qualitativa numa variável quantitativa!

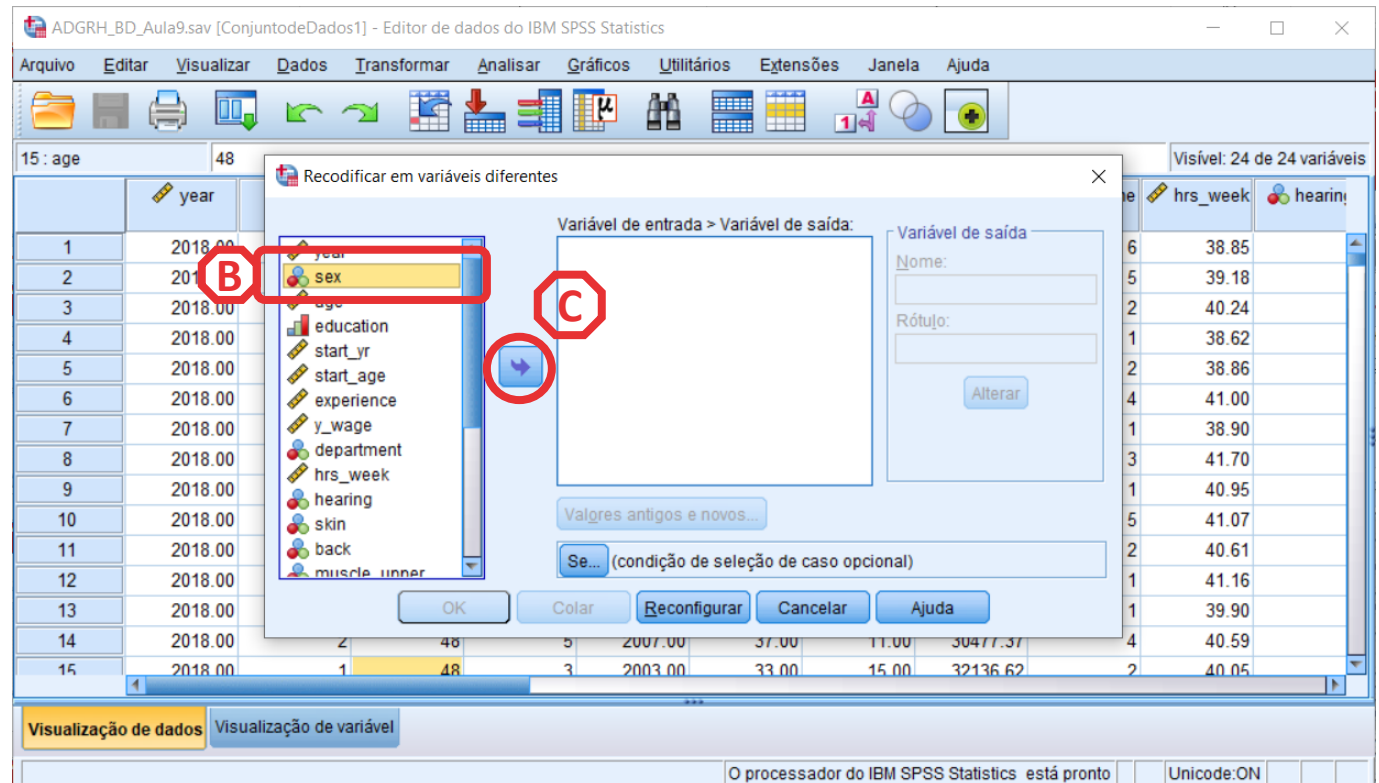
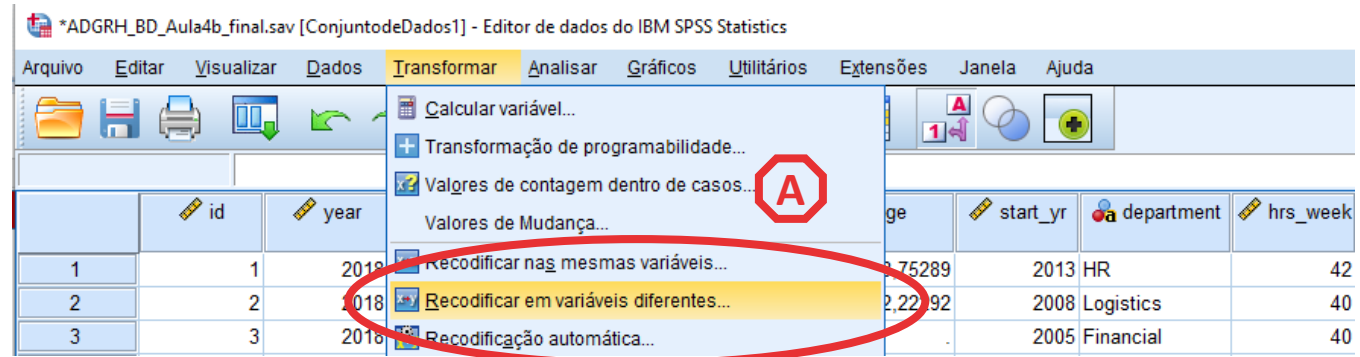
Adicionar IVs nominais

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

A

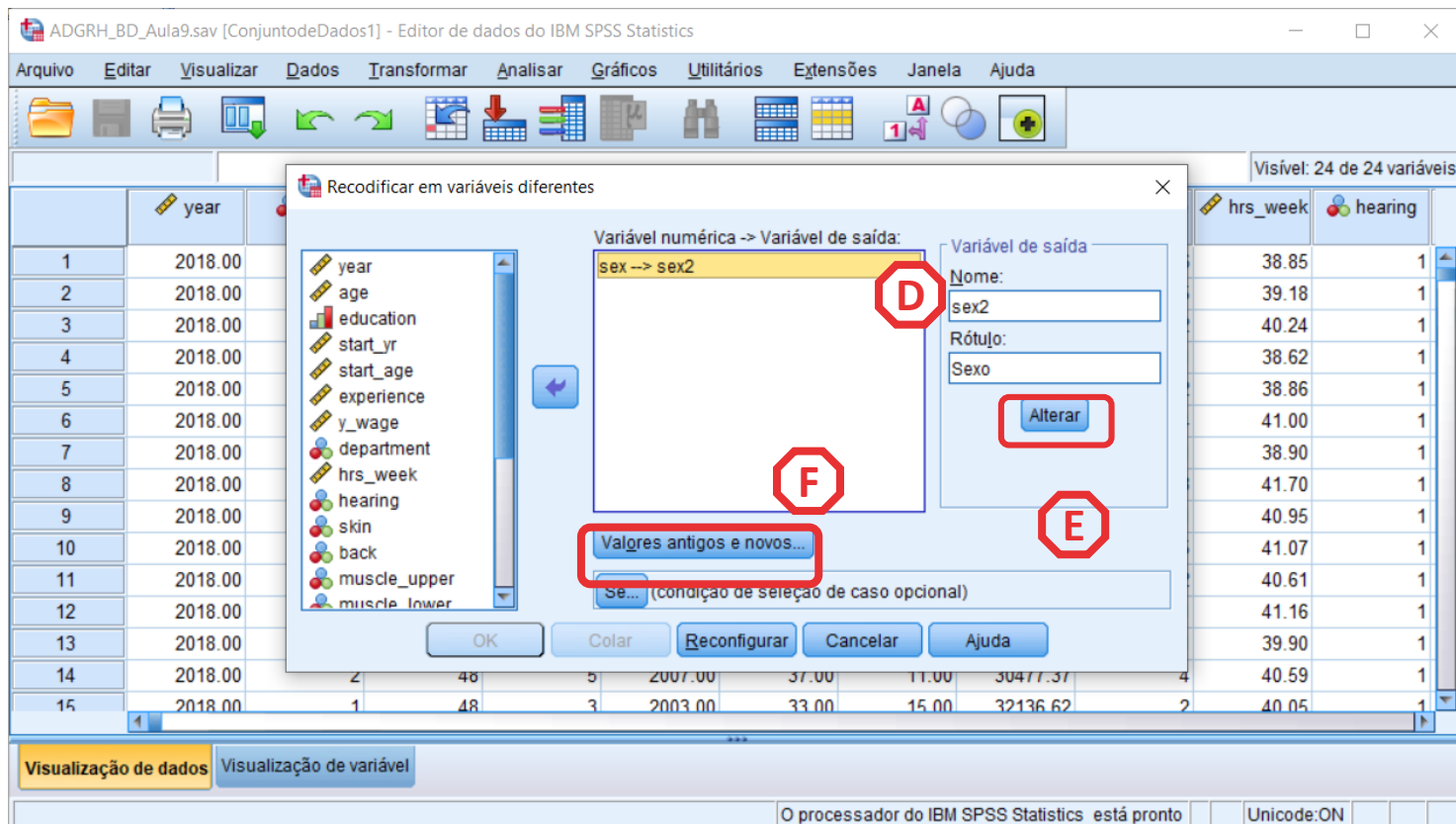
B

C



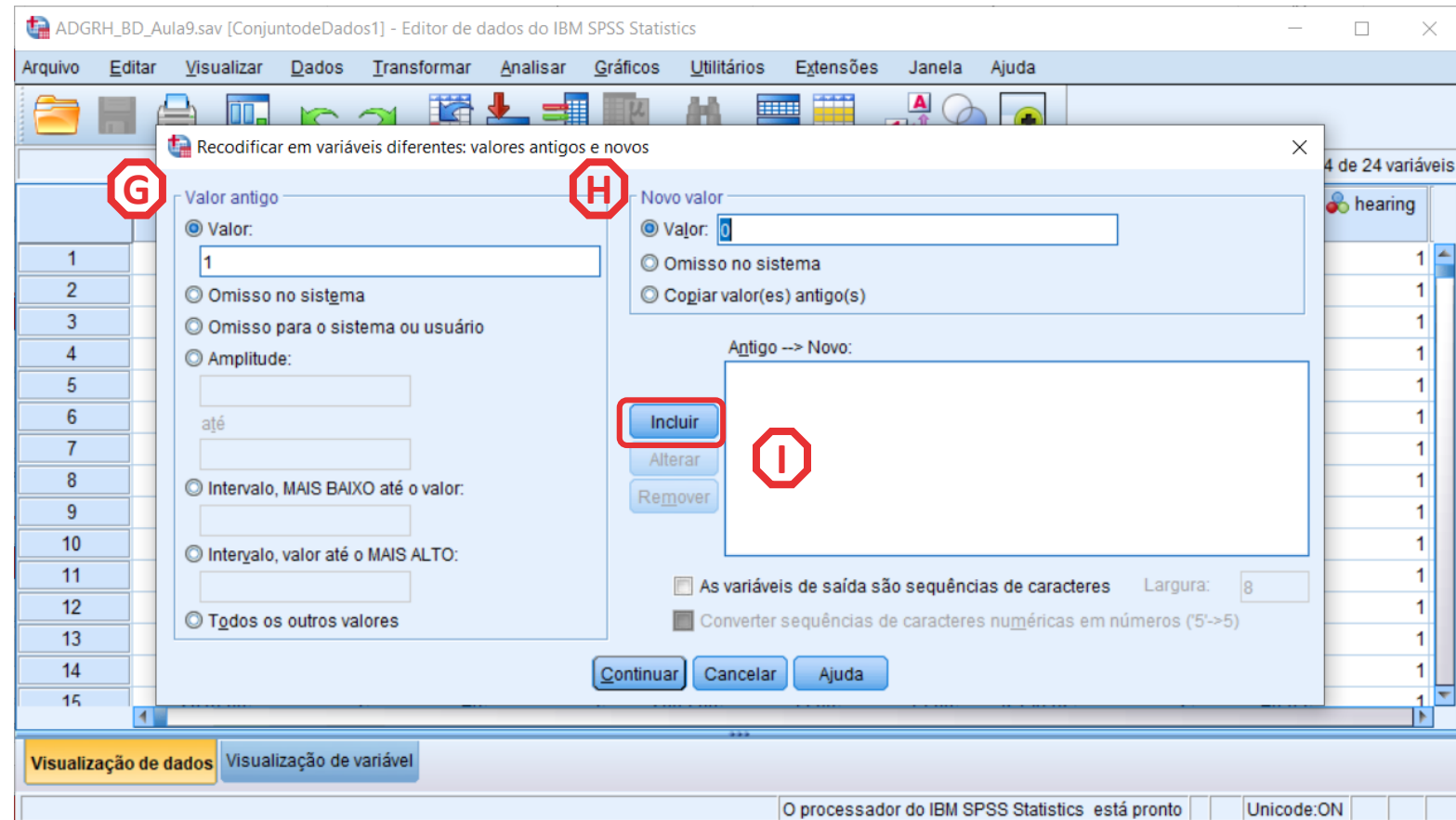
Adicionar IVs nominais

- 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**



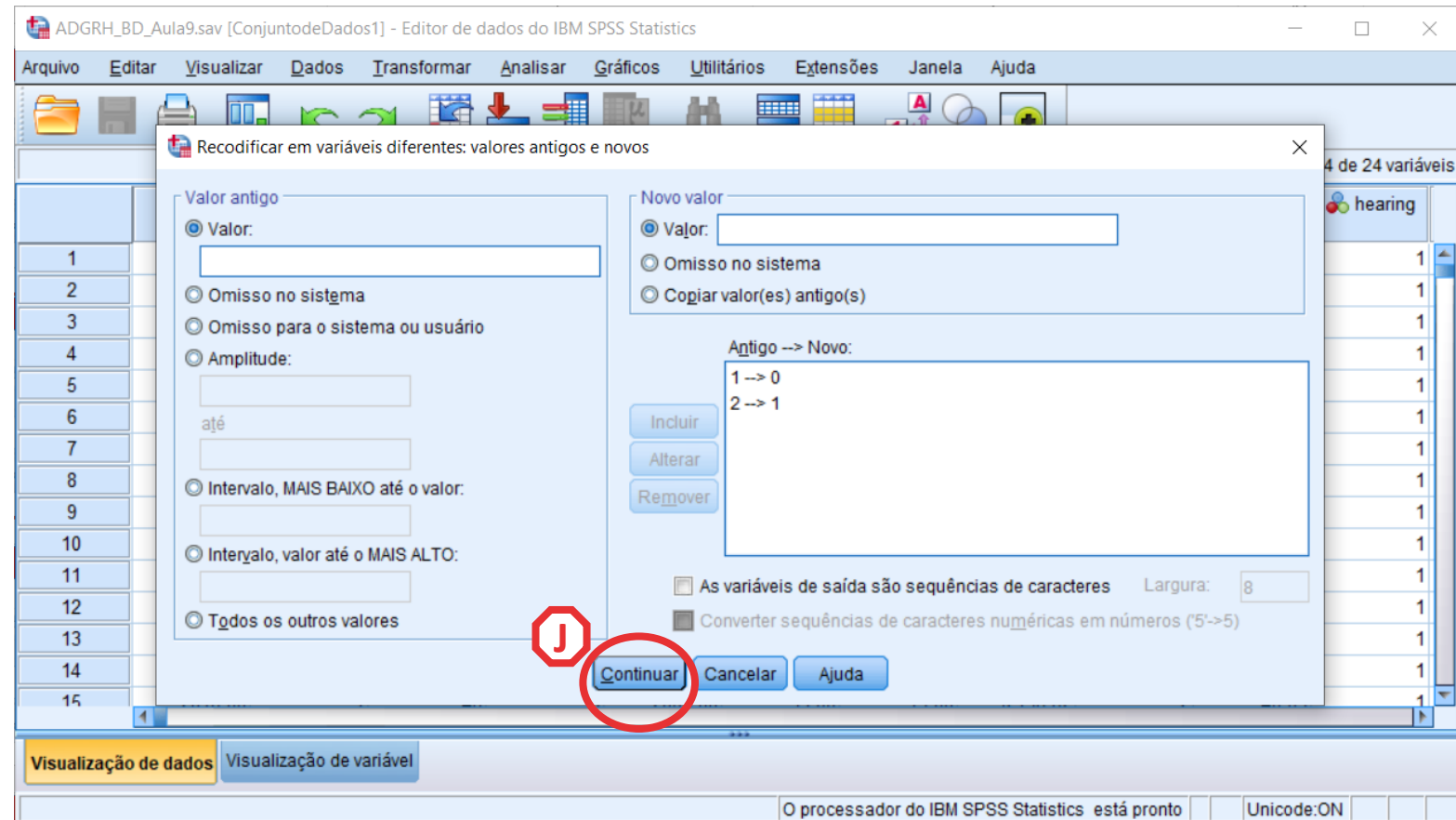
Adicionar IVs nominais

- Vamos definir que ao valor '2' na variável 'sex' Ⓜ
- Corresponde o valor '0', na nova variável ('sex_female') Ⓜ
- Selecionar 'Incluir' Ⓜ



Adicionar IVs nominais

- Vamos definir que ao valor '2' na variável 'sex' G
- Corresponde o valor '0', na nova variável ('sex_female') H
- Selecionar 'Incluir' I
- Selecionar 'Continuar' / 'OK' J



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

Adicionar IVs nominais

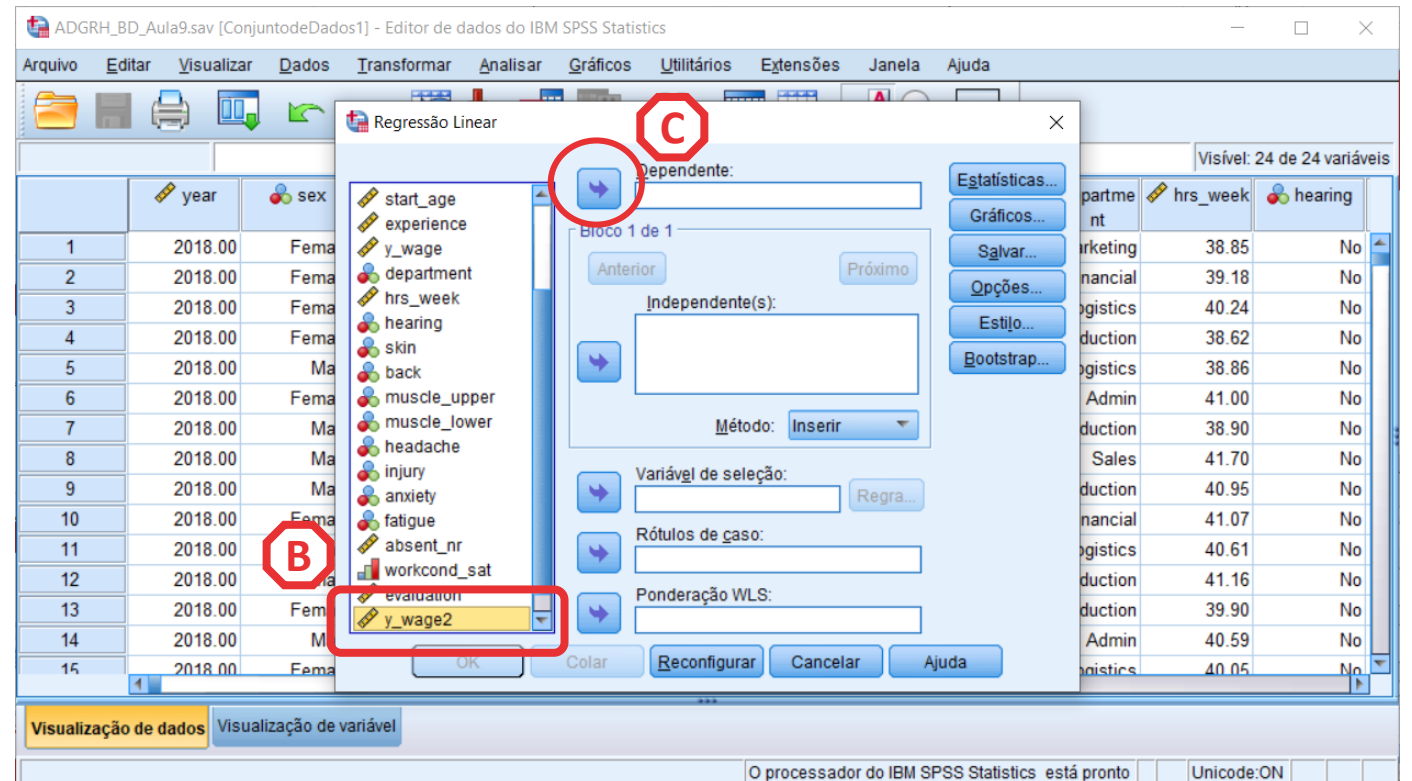
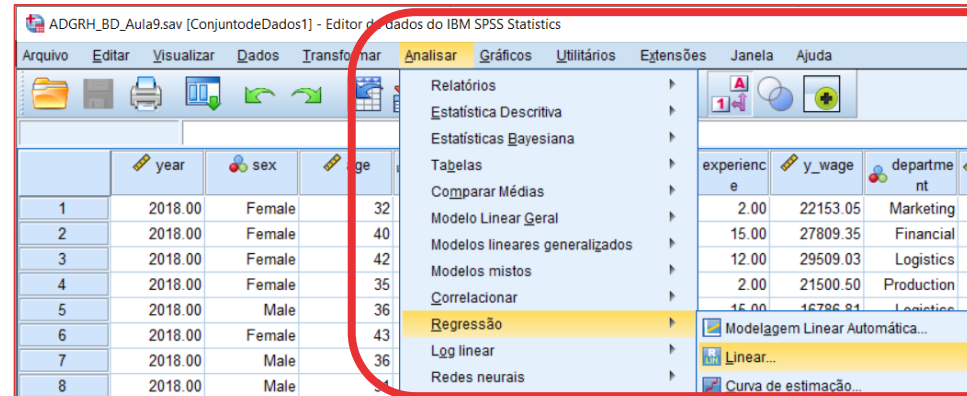
- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y_wage2'
- Colocar na caixa 'Dependente'

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

A

B

C



Adicionar IVs nominais

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

- Selecionar a variável 'y_wage'

- Colocar na caixa 'Dependente'

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

- Selecionar a variável 'sex_female'

- Colocar na caixa 'Independente(s)'

A

B

C

D

D

E

The screenshot shows the 'Regressão Linear' dialog box in IBM SPSS Statistics. The 'Dependente:' field contains 'y_wage2'. The 'Independente(s):' field contains 'experience' and 'evaluation'. The 'Sexo [sex2]' variable is highlighted in the list of independent variables. Red circles and letters A-E mark key steps: A (Analisar/Regressão/Linear), B (y_wage), C (Dependente), D (sex_female), and E (experience/evaluation in Independente(s)).

Adicionar IVs nominais

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

A

- Selecionar a variável 'y_wage'

B

- Colocar na caixa 'Dependente'

C

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

- Selecionar a variável 'sex_female'

D

- Colocar na caixa 'Independente(s)'

E

- Selecionar botão 'Estatísticas'

F

The screenshot shows the IBM SPSS Statistics interface with the 'Regressão Linear' dialog box open. The 'Dependente:' field contains 'y_wage2'. The 'Independente(s):' field contains 'experience', 'evaluation', and 'Sexo [sex2]'. The 'Estatísticas...' button is highlighted with a red box and labeled 'E'. The background shows a data editor with columns 'fatigue' and 'absent'.

Adicionar IVs nominais

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



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

G Coeficientes de regressão

- Estimativas
- Intervalos de confiança
- Matriz de covariâncias

Nível (%): 95

H 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

I Continuar Cancelar Ajuda

OK Colar Reconfigurar Cancelar Ajuda

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 99%

O efeito de ser female é mais relevante que a experiência menor de todas as variáveis



Ser mulher diminui o salário (anual) em -40 Euros, em relação aos homens.



The screenshot shows the SPSS interface with a project tree on the left and a results window on the right. The results window displays the following tables:

Regressão

Variáveis Inseridas/Removidas^a

Modelo	Variáveis inseridas	Variáveis removidas	Método
1	sex_female, experience, evaluation ^b		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 ^a	,997	,997	472,85154

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

ANOVA^a

Modelo		Soma dos Quadrados	df	Quadrado Médio	Z	Sig.
1	Regressão	3,202E+11	3	1,067E+11	477294,848	,000 ^b
	Resíduo	1085298967	4854	223588,580		
	Total	3,212E+11	4857			

a. Variável Dependente: y_wage2
b. Preditores: (Constante), sex_female, experience, evaluation

Coefficientes^a

Modelo		Coefficientes não padronizados		Coefficientes padronizados		Sig.
		B	Erro Erro	Beta	t	
1	(Constante)	-146896,289	148,982		-985,997	,000
	experience	-341,654	1,326	-,240	-257,703	,000
	evaluation	3356,937	2,981	1,079	1126,064	,000
	sex_female	-39,836	14,061	-,002	-2,833	,005

a. Variável Dependente: y_wage2



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
Mas o SPSS agora tem um comando que nos facilita a vida!**

Criar dummies

- Transformar /Criar Dummy **A**
- Selecionar a variável **B**
- Selecionar o pré-fixo **C**

The screenshot shows the SPSS software interface. The 'Transform' menu is open, and the 'Create Dummy Variables' option is highlighted with a red box and labeled 'A'. The 'Create Dummy Variables' dialog box is open, showing the 'education' variable selected in the 'Variables' list, labeled 'B'. The 'Main Effect Dummy Variables' section is checked, and the 'Root Names (One Per Selected Variable)' field contains 'edu_', labeled 'C'. The 'Value Order' is set to 'Ascending'. The 'Measurement Level Usage' section is also visible.

Name	Type	Values	Missing	Columns	Align
1 year	Numeric	{1, No}...	None	8	Right
2 sex	Numeric	{1, Female}...	None	8	Right
3 age	Numeric	{1, No}...	None	8	Right
4 education	Numeric	{1, No}...	None	8	Right
5 start_yr	Numeric	{1, No}...	None	8	Right
6 start_age	Numeric	{1, No}...	None	8	Right
7 experience	Numeric	{1, No}...	None	8	Right
8 y_wage	Numeric	{1, No}...	None	8	Right
9 department	Numeric	{1, No}...	None	8	Right
10 hrs_week	Numeric	{1, No}...	None	8	Right
11 hearing	Numeric	{1, No}...	None	8	Right
12 skin	Numeric	{1, No}...	None	8	Right
13 back	Numeric	{1, No}...	None	8	Right
14 muscle_upper	Numeric	{1, No}...	None	8	Right
15 muscle_lower	Numeric	{1, No}...	None	8	Right
16 headache	Numeric	{1, No}...	None	8	Right
17 injury	Numeric	{1, No}...	None	8	Right
18 anxiety	Numeric	{1, No}...	None	8	Right
19 fatigue	Numeric	{1, No}...	None	8	Right
20 absent_nr	Numeric	{1, No}...	None	8	Right
21 workcond_sat	Numeric	{1, Very sati...}	None	8	Right
22 evaluation	Numeric	{1, No}...	None	8	Right
23 y_wage2	Numeric	{1, No}...	None	8	Right
24 evaluation2	Numeric	{1, No}...	None	8	Right
25 edu_1	Numeric	{1, No}...	None	10	Right
26 edu_2	Numeric	{1, No}...	None	10	Right
27 edu_3	Numeric	{1, No}...	None	10	Right
28 edu_4	Numeric	{1, No}...	None	10	Right
29 edu_5	Numeric	{1, No}...	None	10	Right

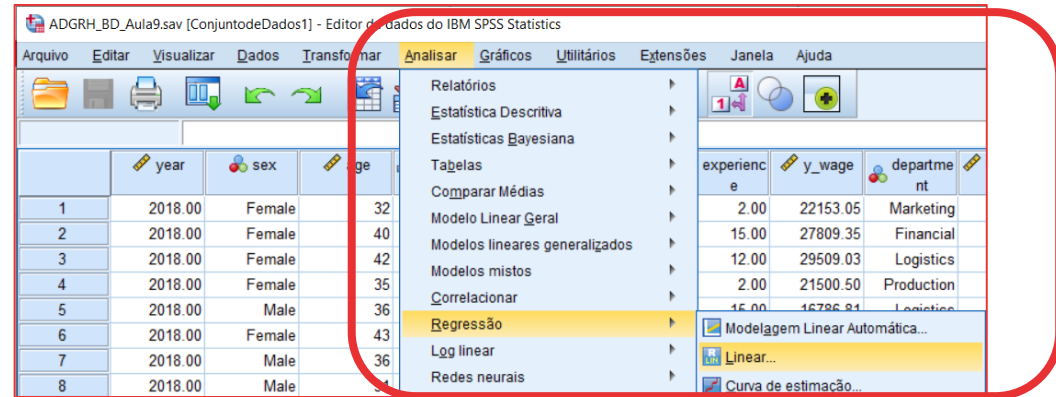
Adicionar IVs nominais

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y_wage'
- Colocar na caixa 'Dependente'
- Colocar na caixa 'Independente'
- Colocar as variáveis 'sex2', 'edu__1', 'edu__2', 'edu__3', 'edu__4', 'edu__5'

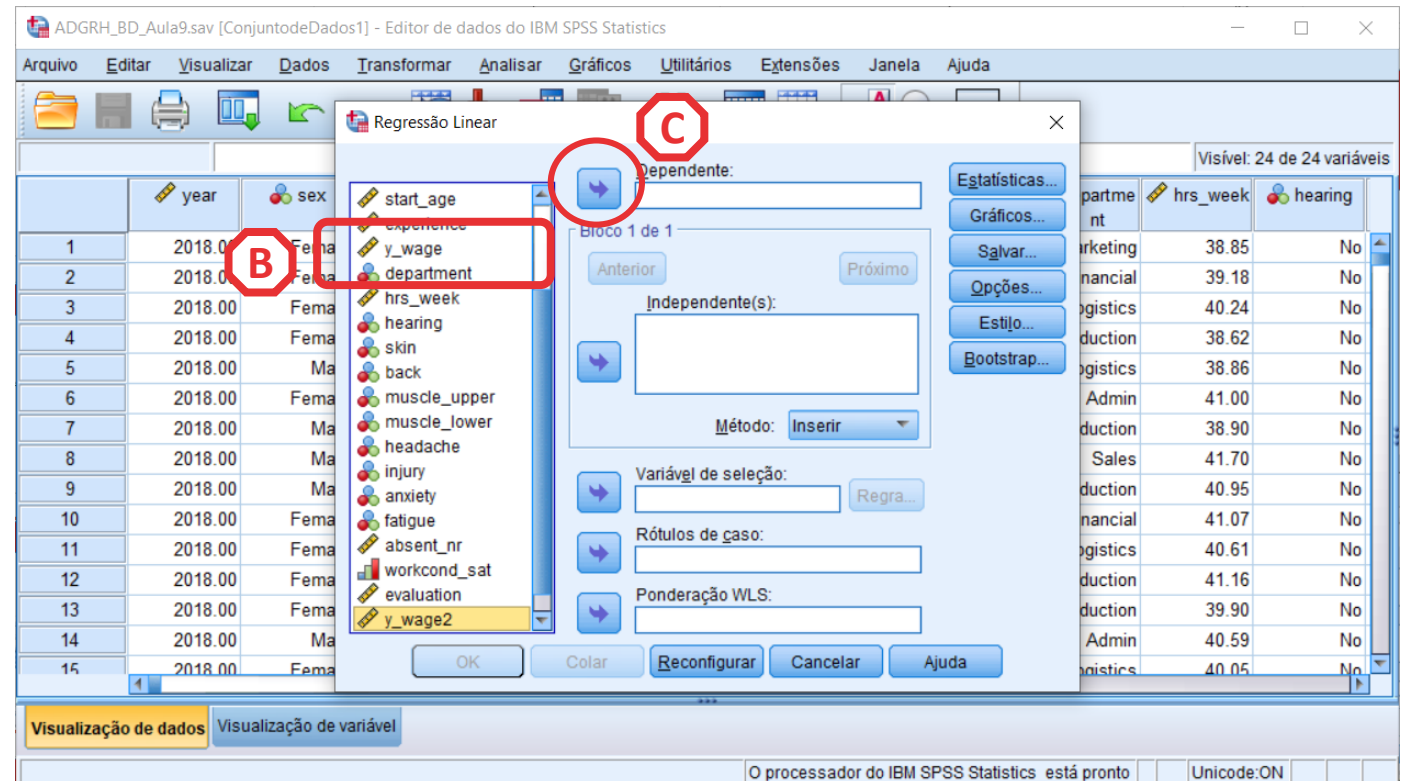
A

B

C



A



Adicionar IVs nominais

- As variáveis categóricas não podem ser incluídas diretamente em uma regressão. As "dummy variables" permitem representar categorias categóricas em formato numérico, com 0s e 1s.
- Numa variável com N categorias, só precisamos de N-1 dummies para diferenciar as N categorias no modelo ("one-hot encoding")
- Para além de ser mais eficiente (e evitar problemas de multicolinearidade), é mais fácil de interpretar: os coeficientes indicam o efeito de cada categoria em relação à categoria de referência (que é representada pelos 0s nas outras "dummy variables").

	X_0	X_1	X_2
Cool	1	0	0
Cooler	0	1	0
Coolest	0	0	1

Adicionar IVs nominais

- Em Estatísticas, selecionar 'Estimativas'
- Selecionar 'Ajuste do modelo'
- Selecionar 'Continuar'/'OK'

D

E

F

The screenshot shows the IBM SPSS Statistics interface with the 'Regressão Linear' dialog box open. The dialog box is titled 'Regressão Linear' and has a 'Dependente:' field. The 'Coeficientes de regressão' section is highlighted with a red box and contains the following options: Estimativas, Intervalos de confiança, and Matriz de covariâncias. The 'Ajuste do modelo' section is also highlighted with a red box and contains the following options: Ajuste do modelo, Alteração de R quadrado, Descritivos, Correlações parciais e de parte, and Diagnósticos de colinearidade. The 'Residuais' section contains the following options: Durbin-Watson, Diagnóstico por caso, Valgres discrepantes (de fora: 3 desvios padrão), and Todos os casos. The 'Continuar' button is circled in red. The 'OK' button is also visible. The background shows a data table with columns like 'year', 'sex', 'hrs_week', and 'hearing'.

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%

Porque é que os coeficientes são negativos?

Coeficientes^a

Modelo		Coeficientes não padronizados		Coeficientes padronizados	t	Sig.
		B	Erro Erro	Beta		
1	(Constante)	-144227,075	34,762		-4148,973	,000
	experience	-331,698	,301	-,233	-1103,779	,000
	evaluation	3316,236	,685	1,065	4842,327	,000
	sex_female	1,635	3,170	,000	,516	,606
	edu_1 education=ISCED 1	-1329,164	4,930	-,065	-269,590	,000
	edu_2 education=ISCED 2	-999,174	4,845	-,049	-206,228	,000
	edu_3 education=ISCED 3	-665,862	4,830	-,033	-137,870	,000
	edu_4 education=ISCED 4	-325,944	4,844	-,016	-67,291	,000

a. Variável Dependente: y_wage2

Adicionar IVs nominais

- O que podemos concluir com esta tabela?
- **Experiência, performance e educação com efeitos relevantes no rendimento**
- Experiência com uma associação negativa com o rendimento
- Performance com uma associação positiva com o rendimento
- O efeito do género não é significativo no modelo **(controlando o efeito da educação, perdeu relevância)**
- As variáveis que avaliam o efeito da educação tomam como categoria de referência a categoria mais elevada ISCED4+. Os coeficientes informam que todas as categorias de educação no modelo se associam a valores de rendimento mais baixos do que o escalão mais qualificado. Educação influencia o rendimento, na medida em que alta educação se associa a valores mais altos de rendimento



Coeficientes^a

Modelo		Coeficientes não padronizados		Coeficientes padronizados	t	Sig.
		B	Erro	Beta		
1	(Constante)	-144227,075	34,762		-4148,973	,000
	experience	-.331,698	,301	-.233	-1103,779	,000
	evaluation	.3316,236	,685	1,065	4842,327	,000
	sex_female	1,635	3,170	,000	,516	,606
	edu_1 education=ISCED 1	-1329,164	4,930	-.065	-269,590	,000
	edu_2 education=ISCED 2	-999,174	4,845	-.049	-206,228	,000
	edu_3 education=ISCED 3	-665,862	4,830	-.033	-137,870	,000
	edu_4 education=ISCED 4	-325,944	4,844	-.016	-67,291	,000

a. Variável Dependente: y_wage2

- A Tabela ‘Coeficientes’

Os coeficientes de regressão num modelo de regressão múltipla dizem-no o quanto uma variável independente (VI) específica está associada a mudanças na variável dependente (VD), mantendo todas as outras variáveis independentes constantes. O modelo é ajustado ao impacto das demais variáveis no modelo.

A ter atenção

- Sinal (positivo/negativo)
- Magnitude (valor do coeficiente)
- Significância Estatística
- Coeficientes Padronizados: ~comparar o impacto

Coeficientes^a

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	,261	,548		2,768	,022
Age			,937	,433		2,187	,056

a. Variável Dependente: Weight

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

Método	Definição
Inserir	<ul style="list-style-type: none"> Todas as variáveis são inseridas ao mesmo tempo
Remover	<ul style="list-style-type: none"> Todas as variáveis são removidas num único passo.
Backward (Backward elimination)	<ul style="list-style-type: none"> Procedimento de seleção em que todas as variáveis são inseridas na equação e removidas sequencialmente. 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. O procedimento pára quando não há variáveis na equação que satisfaçam os critérios de remoção.
Redireccionar (forward selection) 😊	<ul style="list-style-type: none"> Procedimento de seleção no qual as variáveis são inseridas sequencialmente no modelo. 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. O procedimento pára quando houver variáveis que atendam ao critério de entrada.
Stepwise 😞	<ul style="list-style-type: none"> 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. Variáveis já na equação de regressão são removidos se sua probabilidade de F se tornar suficientemente grande. O método pára quando não houver mais variáveis elegíveis para inclusão ou remoção.

Reportar os resultados de um correlação

Reportar os resultados do *Shapiro-Wilk*

Reportar os resultados do *Teste t*

Reportar os resultados do *Teste U*

Reportar os resultados da ANOVA

Reportar os resultados da Kruskal-Wallis

Reportar os resultados do Qui-Quadrado

Reportar os resultados do regressão

Reportar os resultados de um correlação

Quanto à relação entre a variável "Idade" e a variável "Felicidade no Trabalho," observa-se uma correlação negativa estatisticamente significativa ($r = -0.257$, $p < 0.05$). Embora a força dessa associação seja fraca, os resultados sugerem que o aumento de idade está associado à diminuição na felicidade no trabalho.

Reportar os resultados do Shapiro-Wilk

Com base nos resultados do teste de Shapiro-Wilk ($W_{(4858)} = 0.999$, $p = 0.124$), podemos concluir que não há evidências estatisticamente significativas para rejeitar a hipótese nula de que a distribuição da amostra segue uma distribuição normal.

Reportar os resultados do Teste t

Com base nos resultados do teste t de Student ($t(58) = 2.15$, $p = 0.01$), podemos concluir que a diferença entre o grupo de funcionários do Departamento A (Média = 45.2, Desvio Padrão = 6.8) e do Departamento B (Média = 42.8, Desvio Padrão = 7.1) é estatisticamente significativa ($p < 0.05$).

Reportar os resultados do Teste U

Para comparar o nível de escolaridade entre os dois grupos, optou-se por usar o Teste U de Mann-Whitney. Com base nos resultados, podemos concluir que os departamentos diferem na distribuição pelos níveis de escolaridade ISCED7 ($U = 1872$, $p = 0.035$). O Departamento A (Mediana = 5) demonstra, em geral, um nível de escolaridade ISCED7 mais alto do que o Departamento B (Mediana = 4).

Reportar os resultados da Kruskal Wallis

A análise de Kruskal-Wallis revelou uma diferença globalmente significativa entre os três grupos nos resultados da escala de Felicidade no Trabalho ($H(2) = 7.42, p = 0.02$). Posteriormente, foram conduzidos testes de Dunn-Bonferroni para identificar quais grupos diferem estatisticamente entre si. Os testes de Dunn-Bonferroni identificaram diferenças significativas apenas entre os grupos A e B, indicando que os funcionários do Departamento A têm um desempenho estatisticamente diferente em relação à Felicidade no Trabalho em comparação com os do Departamento B.

Reportar os resultados da ANOVA

A análise de variância (ANOVA) revelou uma diferença globalmente significativa entre os três grupos nos resultados da escala Felicidade no trabalho ($F(2, 57) = 4.12, p = 0.02$). Posteriormente, foram realizados testes de Bonferroni para comparar as médias dos grupos a um nível de confiança de 95%. Os testes de Bonferroni identificaram diferenças significativas apenas entre os grupos A (Média = 45.2, Desvio Padrão = 6.8) e B (Média = 42.8, Desvio Padrão = 7.1), demonstrando que os funcionários do Departamento A apresentam uma média estatisticamente superior aos dos Departamento B.

Reportar os resultados da Regressão

Procurou-se compreender o efeito das variáveis experiência e desempenho no rendimento, independentemente do gênero e do nível de educação. A análise de regressão linear revelou um modelo altamente significativo ($F(7, 4850) = 2504.480, p < 0.001$), indicando que várias variáveis independentes têm efeitos significativos na variável dependente ($R^2 = 0.78$). A variável de desempenho exibiu uma associação positiva significativa ($\beta = 1,07, p < 0.001$) com o rendimento. Um aumento de 1 ponto no desempenho está associado a um aumento de €3316 no rendimento, em média. A variável de experiência demonstra uma associação negativa significativa ($\beta = -0.23, p < 0.001$) com o rendimento, ainda que menos relevante. No que diz respeito às variáveis de controle, observa-se que o gênero não se associada de forma significativa ao rendimento ($p > 0.05$). Em contrapartida, as variáveis relacionadas com a educação t mostraram associações significativas com o rendimento. Em relação ao nível de escolaridade mais elevando (categoria de referência omitido na regressão), os restantes níveis de escolaridade estão associados valores de rendimento estatisticamente inferiores, pelo que se constata pelos coeficientes negativos no modelo.