

Aula 7:

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

Validação de Modelos de Regressão Linear

Docente: Daniela Craveiro  
dcraveiro@iseg.ulisboa.pt

- **Na Aula Anterior**
  - **Aprendemos a implementar e interpretar o resultado de um modelo de regressão linear**
- **Objetivos da Aula**
  - **Parte Teórica**
    - **Perceber qual a necessidade de fazermos diagnósticos aos pressupostos do nosso modelo de regressão**
    - **Saber quais são os pressupostos do modelo de regressão linear**
    - **Saber como, com a ajuda de gráficos e testes estatísticos, podemos conferir se os pressupostos do modelo estão a ser cumpridos**
  - **Parte Prática**
    - **Saber implementar os estudo dos pressupostos do modelo no SPSS**

## Pressupostos do modelo de regressão

	<b>PRESSUPOSTOS</b>	<b>DEFINIÇÃO</b>	<b>FORMA DE VALIDAÇÃO</b>
I	<b>Linearidade</b>	<b>O efeito das variáveis independentes na variável dependente é linear e aditivo.</b>	<ul style="list-style-type: none"><li><b>Análise gráfica (Matriz de Dispersão, por exemplo)</b></li></ul>

## Pressupostos do modelo de regressão

	<b>PRESSUPOSTOS</b>	<b>DEFINIÇÃO</b>	<b>FORMA DE VALIDAÇÃO</b>
I	<b>Linearidade</b>	O efeito das variáveis independentes na variável dependente é linear e aditivo.	<ul style="list-style-type: none"><li>• Análise gráfica (Matriz de Dispersão, por exemplo)</li></ul>
II	<b>Normalidade da Distribuição dos Erros</b>	Os erros seguem uma distribuição normal.	<ul style="list-style-type: none"><li>• Análise de Resíduos</li><li>• Gráfico de Q-Q</li></ul>

## Pressupostos do modelo de regressão

	<b>PRESSUPOSTOS</b>	<b>DEFINIÇÃO</b>	<b>FORMA DE VALIDAÇÃO</b>
I	<b>Linearidade</b>	<b>O efeito das variáveis independentes na variável dependente é linear e aditivo.</b>	<ul style="list-style-type: none"><li>• <b>Análise gráfica (Matriz de Dispersão, por exemplo)</b></li></ul>
II	<b>Normalidade da Distribuição dos Erros</b>	<b>Os erros seguem uma distribuição normal.</b>	<ul style="list-style-type: none"><li>• <b>Análise de Resíduos</b></li><li>• <b>Gráfico de Q-Q</b></li></ul>
III	<b>Média Condicional Zero dos Erros</b>	<b>O termo de erro aleatório tem valor esperado igual a zero.</b>	<ul style="list-style-type: none"><li>• <b>Análise de Resíduos</b></li></ul>

## Pressupostos do modelo de regressão

	<b>PRESSUPOSTOS</b>	<b>DEFINIÇÃO</b>	<b>FORMA DE VALIDAÇÃO</b>
I	<b>Linearidade</b>	O efeito das variáveis independentes na variável dependente é linear e aditivo.	<ul style="list-style-type: none"><li>• Análise gráfica (Matriz de Dispersão, por exemplo)</li></ul>
II	<b>Normalidade da Distribuição dos Erros</b>	Os erros seguem uma distribuição normal.	<ul style="list-style-type: none"><li>• Análise de Resíduos</li><li>• Gráfico de Q-Q</li></ul>
III	<b>Média Condicional Zero dos Erros</b>	O termo de erro aleatório tem valor esperado igual a zero.	<ul style="list-style-type: none"><li>• Análise de Resíduos</li></ul>
IV	<b>Homocedasticidade (ou Igual Variância)</b>	A distribuição dos erros apresenta uma variância constante (hipótese da homocedasticidade).	<ul style="list-style-type: none"><li>• Análise de Resíduos</li></ul>

## Pressupostos do modelo de regressão

	<b>PRESSUPOSTOS</b>	<b>DEFINIÇÃO</b>	<b>FORMA DE VALIDAÇÃO</b>
I	<b>Linearidade</b>	O efeito das variáveis independentes na variável dependente é linear e aditivo.	<ul style="list-style-type: none"><li>• Análise gráfica (Matriz de Dispersão, por exemplo)</li></ul>
II	<b>Normalidade da Distribuição dos Erros</b>	Os erros seguem uma distribuição normal.	<ul style="list-style-type: none"><li>• Análise de Resíduos</li><li>• Gráfico de Q-Q</li></ul>
III	<b>Média Condicional Zero dos Erros</b>	O termo de erro aleatório tem valor esperado igual a zero.	<ul style="list-style-type: none"><li>• Análise de Resíduos</li></ul>
IV	<b>Homocedasticidade (ou Igual Variância)</b>	A distribuição dos erros apresenta uma variância constante (hipótese da homocedasticidade).	<ul style="list-style-type: none"><li>• Análise de Resíduos</li></ul>
V	<b>Independência dos Erros</b>	Os erros não estão correlacionados, i.e., o valor de um erro não depende de qualquer outro erro.	<ul style="list-style-type: none"><li>• Dublin-Watson</li></ul>

## Pressupostos do modelo de regressão

	<b>PRESSUPOSTOS</b>	<b>DEFINIÇÃO</b>	<b>FORMA DE VALIDAÇÃO</b>
I	<b>Linearidade</b>	O efeito das variáveis independentes na variável dependente é linear e aditivo.	<ul style="list-style-type: none"> <li>• Análise gráfica (Matriz de Dispersão, por exemplo)</li> </ul>
II	<b>Normalidade da Distribuição dos Erros</b>	Os erros seguem uma distribuição normal.	<ul style="list-style-type: none"> <li>• Análise de Resíduos</li> <li>• Gráfico de Q-Q</li> </ul>
III	<b>Média Condicional Zero dos Erros</b>	O termo de erro aleatório tem valor esperado igual a zero.	<ul style="list-style-type: none"> <li>• Análise de Resíduos</li> </ul>
IV	<b>Homocedasticidade (ou Igual Variância)</b>	A distribuição dos erros apresenta uma variância constante (hipótese da homocedasticidade).	<ul style="list-style-type: none"> <li>• Análise de Resíduos</li> </ul>
V	<b>Independência dos Erros</b>	Os erros não estão correlacionados, i.e., o valor de um erro não depende de qualquer outro erro.	<ul style="list-style-type: none"> <li>• Dublin-Watson</li> </ul>
VI	<b>Ausência de multicolinearidade perfeita</b>	As variáveis independentes não estão perfeitamente correlacionadas entre si.	<ul style="list-style-type: none"> <li>• Diagnósticos de Colinearidade</li> </ul>



## Pressupostos do modelo de regressão

	<b>PRESSUPOSTOS</b>	<b>DEFINIÇÃO</b>	<b>FORMA DE VALIDAÇÃO</b>
I	<b>Linearidade</b>	O efeito das variáveis independentes na variável dependente é linear e aditivo.	<ul style="list-style-type: none"> <li>• Análise gráfica (Matriz de Dispersão, por exemplo)</li> </ul>
II	<b>Normalidade da Distribuição dos Erros</b>	Os erros seguem uma distribuição normal.	<ul style="list-style-type: none"> <li>• Análise de Resíduos</li> <li>• Gráfico de Q-Q</li> </ul>
III	<b>Média Condicional Zero dos Erros</b>	O termo de erro aleatório tem valor esperado igual a zero.	<ul style="list-style-type: none"> <li>• Análise de Resíduos</li> </ul>
IV	<b>Homocedasticidade (ou Igual Variância)</b>	A distribuição dos erros apresenta uma variância constante (hipótese da homocedasticidade).	<ul style="list-style-type: none"> <li>• Análise de Resíduos</li> </ul>
V	<b>Independência dos Erros</b>	Os erros não estão correlacionados, i.e., o valor de um erro não depende de qualquer outro erro.	<ul style="list-style-type: none"> <li>• Dublin-Watson</li> </ul>
VI	<b>Ausência de multicolinearidade perfeita</b>	As variáveis independentes não estão perfeitamente correlacionadas entre si.	<ul style="list-style-type: none"> <li>• Diagnósticos de Colinearidade</li> </ul>
VII	<b>Ausência de Observações Influentes</b>	Não existem observações que tenham uma influência anormal nos resultados do modelo.	<ul style="list-style-type: none"> <li>• Cook's Distance</li> </ul>

## **E qual é o problema se estes pressupostos não se verificarem?**

- **Os intervalos de confiança ou os p-values podem estar a ser subestimados (i.e. mais pequenos do que na realidade são) ...**

**ou seja: estamos a atribuir significância estatística a uma estimativa que na realidade não a terá!**

**Como podemos saber se estes pressupostos estão a ser cumpridos?**

## Validação do Modelo de Regressão Linear

- 1. Estimar o modelo de regressão com os diagnósticos*
- 2. Avaliação do Pressuposto II: Normalidade da Distribuição dos Erros*
- 3. Avaliação do Pressuposto III: Média Condicional Zero dos Erros*
- 4. Avaliação do Pressuposto IV: Homocedasticidade*
- 4. Avaliação do Pressuposto V: Independência dos Erros*
- 5. Avaliação do Pressuposto VI: Ausência de Multicolinearidade Perfeita*
- 6. Avaliação do Pressuposto VII: Ausência de Observações Influentes*

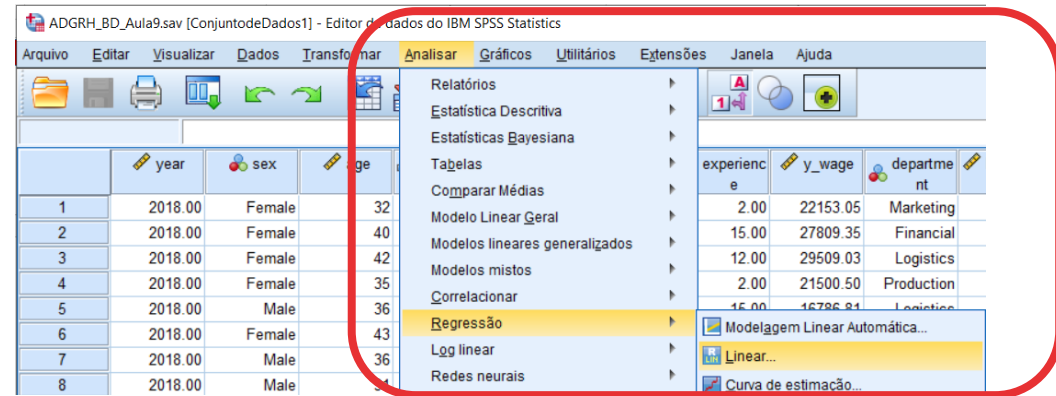
## Validação do Modelo de Regressão Linear

*1. Estimar o modelo de regressão com os diagnósticos*

# Diagnósticos

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

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

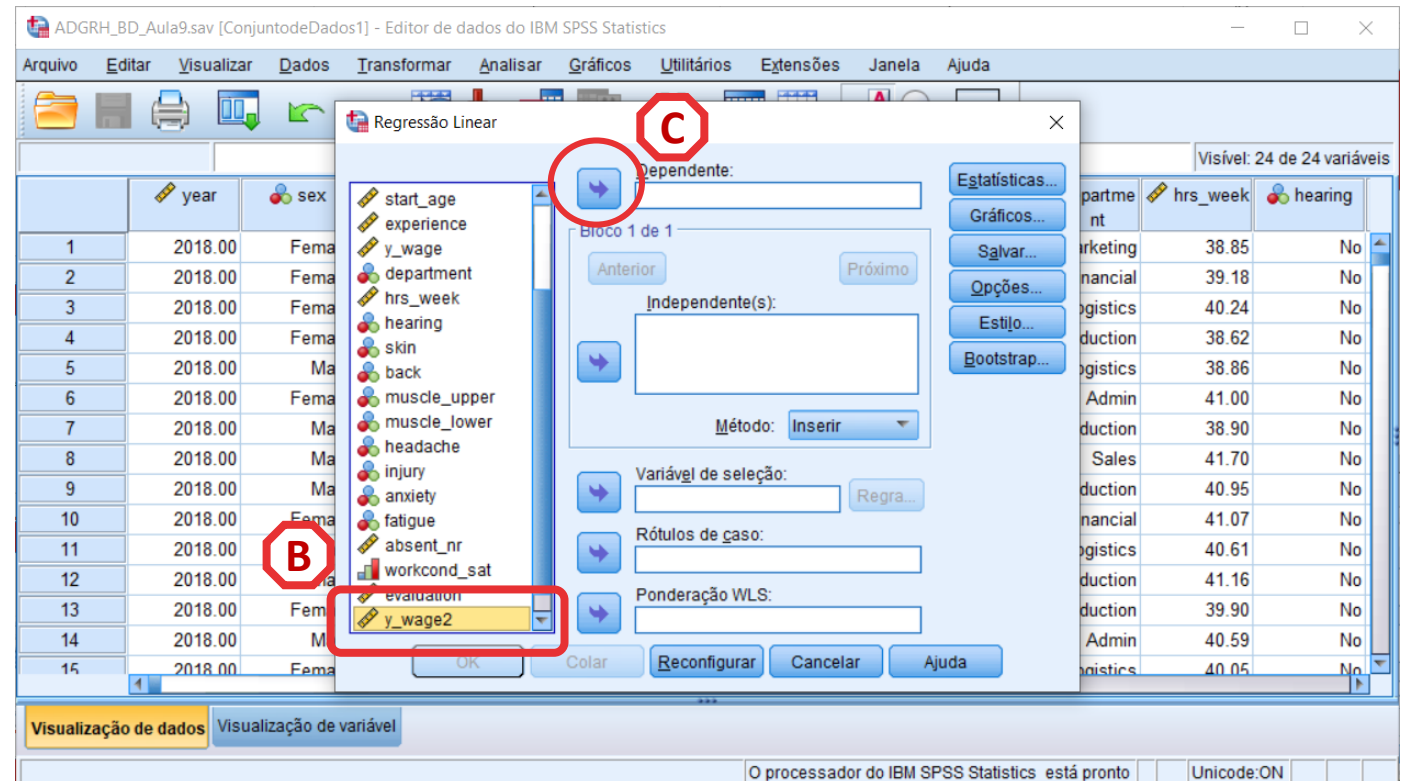


A

A

B

C



B

C

# Diagnósticos

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

- Colocar na caixa 'Dependente'

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

- Selecionar botão 'Estatísticas'

A

B

C

D

ADGRH\_BD\_Aula10.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): sex, education, experience

Método: Inserir

Variável de seleção: Regra...

Rótulos de caso:

Ponderação WLS:

OK Color Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

	year	sex	year	sex	hrs_week	hearing
1	2018.00	Fema	2018.00	Fema	38.85	No
2	2018.00	Fema	2018.00	Fema	39.18	No
3	2018.00	Fema	2018.00	Fema	40.24	No
4	2018.00	Fema	2018.00	Fema	38.62	No
5	2018.00	Ma	2018.00	Ma	38.86	No
6	2018.00	Fema	2018.00	Fema	41.00	No
7	2018.00	Ma	2018.00	Ma	38.90	No
8	2018.00	Ma	2018.00	Ma	41.70	No
9	2018.00	Ma	2018.00	Ma	40.95	No
10	2018.00	Fema	2018.00	Fema	41.07	No
11	2018.00	Fema	2018.00	Fema	40.61	No
12	2018.00	Fema	2018.00	Fema	41.16	No
13	2018.00	Fema	2018.00	Fema	39.90	No
14	2018.00	Ma	2018.00	Ma	40.59	No
15	2018.00	Fema	2018.00	Fema	40.05	No

# Diagnósticos

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

- Selecionar a variável 'y\_wage2'

- Colocar na caixa 'Dependente'

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

- Selecionar botão 'Estatísticas'

- Selecionar 'Estimativas'

A

B

C

D

E

The screenshot displays the IBM SPSS Statistics interface. The main window shows a data table with columns: year, sex, education, start, experience, y\_wage2, hrs\_week, and hearing. The 'Regressão Linear' dialog box is open, with the 'Dependente:' field set to 'y\_wage2'. The 'Estatísticas' sub-dialog is also open, showing the following options:

- Coeficientes de regressão
  - 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

The 'Residuais' section shows:

- Durbin-Watson
- Diagnóstico por caso
- Valores discrepantes no lado de fora: 3 desvios padrão
- Todos os casos

Buttons at the bottom of the dialog include 'Continuar', 'Cancelar', and 'Ajuda'. The status bar at the bottom of the SPSS window indicates 'O processador do IBM SPSS Statistics está pronto' and 'Unicode:ON'.



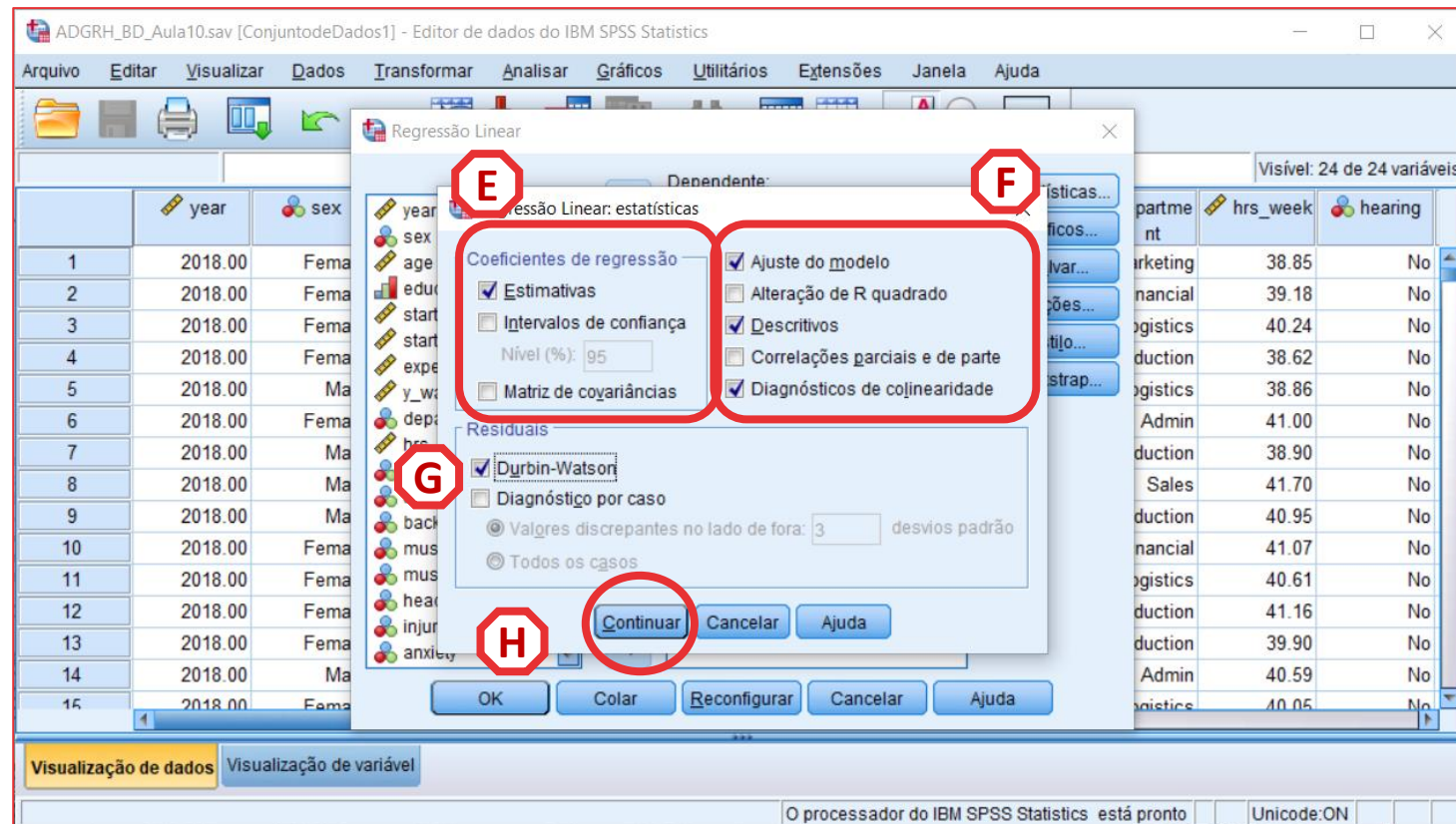
### Diagnósticos

- Selecionar 'Ajuste do modelo'
- Selecionar 'Descritivos'
- Selecionar 'Diagnósticos de colinearidade'
- Selecionar 'Dublin-Watson'
- Selecionar 'Continuar'

F

G

H



# Diagnósticos

- Selecionar botão 'Salvar'



The screenshot displays the IBM SPSS Statistics interface. The main window title is 'ADGRH\_BD\_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics'. The menu bar includes 'Arquivo', 'Editar', 'Visualizar', 'Dados', 'Transformar', 'Analisar', 'Gráficos', 'Utilitários', 'Extensões', 'Janela', and 'Ajuda'. The 'Regressão Linear' dialog box is open, showing the following configuration:

- Dependente:** y\_wage2
- Bloco 1 de 1:** Anterior, Próximo
- Independente(s):** sex, education, experience
- Método:** Inserir
- Variável de seleção:** (empty field), Regra...
- Rótulos de caso:** (empty field)
- Ponderação WLS:** (empty field)

Buttons on the right side of the dialog box include: Estatísticas..., Gráficos..., **Salvar...** (highlighted with a red box and a red octagonal warning icon), Opções..., Estilo..., and Bootstrap... At the bottom of the dialog are buttons for OK, Color, Reconfigurar, Cancelar, and Ajuda. The background data table shows columns for 'year', 'sex', 'y\_wage2', 'hrs\_week', and 'hearing'.

# Diagnósticos

- Selecionar botão 'Salvar'
- Selecionar 'Padronizado'



Regressão Linear: salvar

Valores preditos

- Não padronizado
- Padronizado
- Ajustado
- S.E. de predições médias

Residuais

- Não padronizado
- Padronizado
- Estudentização
- Excluído
- Estudentizado excluído

Distâncias

- Mahalanobis
- de Cook
- Valores de ponto alavanca

Estadísticas de influência

- DfBeta(s)
- DfBeta(s) padronizado(s)
- DfFit
- DfFit padronizado
- Razão de covariância

Intervalos de predição

- Média
- Individual

Intervalo de confiança: 95 %

Estadísticas de coeficiente

- Criar estatísticas de coeficiente
- Criar novo conjunto de dados  
Nome do conjunto de dados:
- Gravar um novo arquivo de dados

Exportar informações do modelo para o arquivo XML

- Incluir a matriz de covariâncias

# Diagnósticos

- Selecionar botão 'Salvar'
- Selecionar 'Padronizado'
- Selecionar 'de Cook' e 'Valores de ponto alavanca'

I

J

K

Regressão Linear: salvar

Valores preditos

- Não padronizado
- Padronizado
- Ajustado
- S.E. de predições médias

Residuais

- Não padronizado
- Padronizado
- Studentização
- Excluído
- Studentizado excluído

Distâncias

- Mahalanobis
- de Cook
- Valores de ponto alavanca

Intervalos de predição

- Média
- Individual

Intervalo de confiança: 95 %

Estadísticas de influência

- DfBeta(s)
- DfBeta(s) padronizado(s)
- DfFit
- DfFit padronizado
- Razão de covariância

Estadísticas de coeficiente

- Criar estatísticas de coeficiente
- Criar novo conjunto de dados  
Nome do conjunto de dados:
- Gravar um novo arquivo de dados

Exportar informações do modelo para o arquivo XML

- Incluir a matriz de covariâncias

# Diagnósticos

- Selecionar botão 'Salvar'
- Selecionar 'Padronizado'
- Selecionar 'de Cook' e 'Valores de ponto alavanca'
- Selecionar 'DfBeta(s) padronizado(s)'
- Selecionar 'Continuar' / 'OK'

I

J

K

L

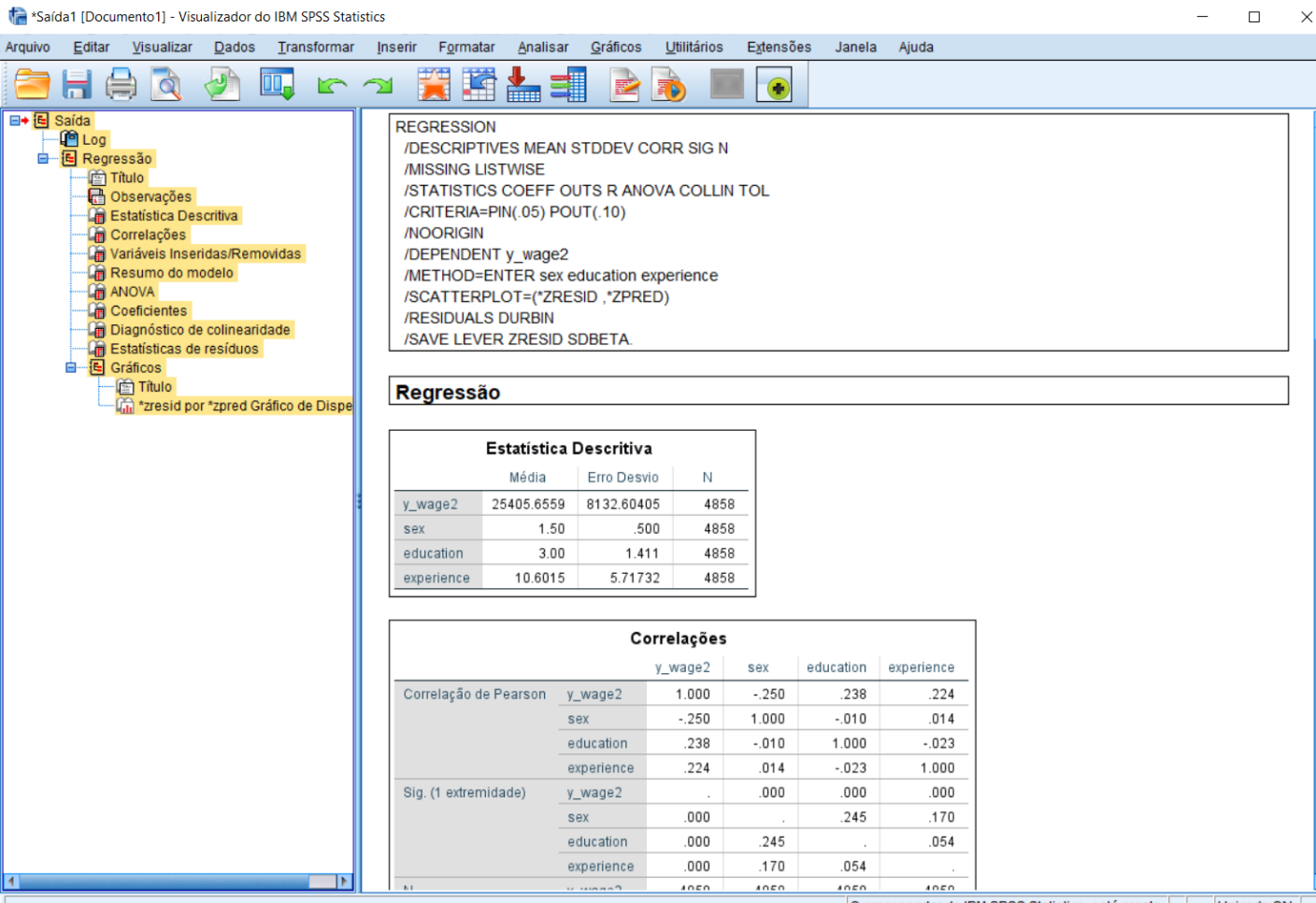
M

The image shows a screenshot of the 'Regressão Linear: salvar' dialog box. The dialog is divided into several sections, with red boxes and letters I-M highlighting specific options:

- Valores preditos:**  Não padronizado,  Padronizado,  Ajustado,  S.E. de predições médias.
- Residuais:**  Não padronizado,  Padronizado,  Estudentização,  Excluído,  Estudentizado excluído.
- Distâncias:**  Mahalanobis,  de Cook,  Valores de ponto alavanca.
- Estadísticas de influência:**  DfBeta(s),  DfBeta(s) padronizado(s),  DfFit,  DfFit padronizado,  Razão de covariância.
- Intervalos de predição:**  Média,  Individual, Intervalo de confiança: 95 %.
- Estadísticas de coeficiente:**  Criar estatísticas de coeficiente,  Criar novo conjunto de dados (Nome do conjunto de dados: ) or  Gravar um novo arquivo de dados (Arquivo...).
- Exportar informações do modelo para o arquivo XML:**  (Navegar...),  Incluir a matriz de covariâncias.
- Buttons:** Continuar (circled in red), Cancelar, Ajuda.

# Diagnósticos

- Os resultados são publicados no 'Visualizador de Resultados'



The screenshot shows the IBM SPSS Statistics Results Viewer window. The left pane displays a tree view of the output, with 'Regressão' expanded to show 'Estadística Descritiva' and 'Correlações'. The main pane displays the regression command and the corresponding statistical results.

```
REGRESSION
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA COLLIN TOL
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT y_wage2
/METHOD=ENTER sex education experience
/SCATTERPLOT=(*ZRESID ,*ZPRED)
/RESIDUALS DURBIN
/SAVE LEVER ZRESID SDBETA.
```

**Regressão**

	Média	Erro Desvio	N
y_wage2	25405.6559	8132.60405	4858
sex	1.50	.500	4858
education	3.00	1.411	4858
experience	10.6015	5.71732	4858

	y_wage2	sex	education	experience
Correlação de Pearson				
y_wage2	1.000	-.250	.238	.224
sex	-.250	1.000	-.010	.014
education	.238	-.010	1.000	-.023
experience	.224	.014	-.023	1.000
Sig. (1 extremidade)				
y_wage2	.	.000	.000	.000
sex	.000	.	.245	.170
education	.000	.245	.	.054
experience	.000	.170	.054	.

# Diagnósticos

- Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis

\*ADGRH\_BD\_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: ZRE\_1 -1.13894522293957 Visível: 31 de 31 variáveis

	evaluation 2	ZRE_1	COO_1	LEV_1	SDB0_1	SDB1_1	SDB2_1	SDB3_1
1	55.17	-1.13895	.00031	.00075	-.02008	.01582	-.01082	.02414
2	51.75	-.73316	.00009	.00043	-.00244	.01050	-.00751	-.00841
3	54.79	.08094	.00000	.00032	.00054	-.00115	.00082	.00032
4	58.30	-.10802	.00000	.00110	-.00361	.00153	.00227	.00237
5	49.89	-1.61943	.00061	.00073	.00170	-.02284	.03244	-.01683
6	52.74	.26865	.00001	.00061	.00066	-.00379	.00543	.00045
7	47.60	-.51252	.00005	.00059	.00244	-.00757	-.00507	.00848
8	52.65	-1.39629	.00044	.00069	-.00962	-.02039	.01471	.02727
9	48.48	-.09011	.00000	.00036	-.00012	-.00130	.00092	.00063
10	52.40	-.09098	.00000	.00061	-.00032	.00128	-.00183	.00008
11	47.21	-.93018	.00018	.00064	-.02087	.01351	.01905	-.00301
12	48.39	.34721	.00004	.00096	.00550	-.00511	-.00699	.00636
13	50.19	1.06687	.00039	.00117	.01455	-.01579	-.02136	.02492
14	40.86	1.27964	.00034	.00063	-.02269	.01872	.02621	.00163
15	45.36	.98336	.00013	.00033	.00845	-.01418	.00006	.01105

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Diagnósticos

- Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis
  - Uma variável com os 'Resíduos Padronizados' da variável dependente (ZRE\_1) para cada observação

\*ADGRH\_BD\_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: ZRE\_1 -1.13894522293957 Visível: 30 de 30 variáveis

	evaluation 2	ZRE_1	LEV_1	SDB0_1	SDB1_1	SDB2_1	SDB3_1	var
1	55.17		5	-.02008	.01582	-.01082	.02414	
2	51.75		3	-.00244	.01050	-.00751	-.00841	
3	54.79		2	.00054	-.00115	.00082	.00032	
4	58.30	-.10802	.00110	-.00361	.00153	.00227	.00237	
5	49.89	-1.61943	.00073	.00170	-.02284	.03244	-.01683	
6	52.74	.26865	.00061	.00066	-.00379	.00543	.00045	
7	47.60	-.51252	.00059	.00244	-.00757	-.00507	.00848	
8	52.65	-1.39629	.00069	-.00962	-.02039	.01471	.02727	
9	48.48	-.09011	.00036	-.00012	-.00130	.00092	.00063	
10	52.40	-.09098	.00061	-.00032	.00128	-.00183	.00008	
11	47.21	-.93018	.00064	-.02087	.01351	.01905	-.00301	
12	48.39	.34721	.00096	.00550	-.00511	-.00699	.00636	
13	50.19	1.06687	.00117	.01455	-.01579	-.02136	.02492	
14	40.86	1.27964	.00063	-.02269	.01872	.02621	.00163	
15	45.36	.98336	.00033	.00845	-.01418	.00006	.01105	

Nome: ZRE\_1  
Rótulo: Standardized Residual  
Tipo: Numérico  
Medida: Escala

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON



## Diagnósticos

- Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis
  - Uma variável com os 'Resíduos Padronizados' da variável dependente (ZRE\_1) para cada observação
  - Uma variável que mede a distancia de Cook associada a cada observação (COO\_1)

\*ADGRH\_BD\_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : COO\_1 .00031099841585 Visível: 31 de 31 variáveis

	evaluation 2	ZRE_1	COO_1	LEV_1	SDB0_1	SDB1_1	SDB2_1	SDB3_1
1	55.17	-1.13895	.00031	.00075	-.02008	.01582	-.01082	.02414
2	51.75	-.73316	.00009	.00043	-.00244	.01050	-.00751	-.00841
3	54.79	.08094	.00000	.00032	.00054	-.00115	.00082	.00032
4	58.30	-.10802	.00000	.00110	-.00361	.00153	.00227	.00237
5	49.89	-1.61943	.00061	.00073	.00170	-.02284	.03244	-.01683
6	52.74	.26865	.00001	.00061	.00066	-.00379	.00543	.00045
7	47.60	-.51252	.00005	.00059	.00244	-.00757	-.00507	.00848
8	52.65	-1.39629	.00044	.00069	-.00962	-.02039	.01471	.02727
9	48.48	-.09011	.00000	.00036	-.00012	-.00130	.00092	.00063
10	52.40	-.09098	.00000	.00061	-.00032	.00128	-.00183	.00008
11	47.21	-.93018	.00018	.00064	-.02087	.01351	.01905	-.00301
12	48.39	.34721	.00004	.00096	.00550	-.00511	-.00699	.00636
13	50.19	1.06687	.00039	.00117	.01455	-.01579	-.02136	.02492
14	40.86	1.27964	.00034	.00063	-.02269	.01872	.02621	.00163
15	45.36	.98336	.00013	.00033	.00845	-.01418	.00006	.01105

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Diagnósticos

- Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis
  - Uma variável com os 'Resíduos Padronizados' da variável dependente (ZRE\_1) para cada observação
  - Uma variável que mede a distancia de Cook associada a cada observação (COO\_1)
  - Uma variável que mede influência relativa de cada observação no ajuste do modelo (LEV\_1).

\*ADGRH\_BD\_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : LEV\_1 .00075130556031 Visível: 31 de 31 variáveis

	evaluation 2	ZRE_1	COO_1	LEV_1	SDB0_1	SDB1_1	SDB2_1	SDB3_1
1	55.17	-1.13895	.00031	.00075	-.02008	.01582	-.01082	.02414
2	51.75	-.73316	.00009	.00043	-.00244	.01050	-.00751	-.00841
3	54.79	.08094	.00000	.00032	.00054	-.00115	.00082	.00032
4	58.30	-.10802	.00000	.00110	-.00361	.00153	.00227	.00237
5	49.89	-1.61943	.00061	.00073	.00170	-.02284	.03244	-.01683
6	52.74	.26865	.00001	.00061	.00066	-.00379	.00543	.00045
7	47.60	-.51252	.00005	.00059	.00244	-.00757	-.00507	.00848
8	52.65	-1.39629	.00044	.00069	-.00962	-.02039	.01471	.02727
9	48.48	-.09011	.00000	.00036	-.00012	-.00130	.00092	.00063
10	52.40	-.09098	.00000	.00061	-.00032	.00128	-.00183	.00008
11	47.21	-.93018	.00018	.00064	-.02087	.01351	.01905	-.00301
12	48.39	.34721	.00004	.00096	.00550	-.00511	-.00699	.00636
13	50.19	1.06687	.00039	.00117	.01455	-.01579	-.02136	.02492
14	40.86	1.27964	.00034	.00063	-.02269	.01872	.02621	.00163
15	45.36	.98336	.00013	.00033	.00845	-.01418	.00006	.01105

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

## Diagnósticos

- Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis
  - Uma variável com os 'Resíduos Padronizados' da variável dependente (ZRE\_1) para cada observação
  - Uma variável que mede a distancia de Cook associada a cada observação (COO\_1)
  - Uma variável que mede influência relativa de cada observação no ajuste do modelo (LEV\_1).
  - Por cada variável independente é criada uma variável com os DFBETA Padronizado, mede a influência de uma dada observação na estimação dos parâmetros.

\*ADGRH\_BD\_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : SDB1\_1 .01582242634099 Visível: 31 de 31 variáveis

	evaluation 2	ZRE_1	COO_1	LEV_1	SDB0_1	SDB1_1	SDB2_1	SDB3_1
1	55.17	-1.13895	.00031	.00075	-.02008	.01582	-.01082	.02414
2	51.75	-.73316	.00009	.00043	-.00244	.01050	-.00751	-.00841
3	54.79	.08094	.00000	.00032	.00054	-.00115	.00082	.00032
4	58.30	-.10802	.00000	.00110	-.00361	.00153	.00227	.00237
5	49.89	-1.61943	.00061	.00073	.00170	-.02284	.03244	-.01683
6	52.74	.26865	.00001	.00061	.00066	-.00379	.00543	.00045
7	47.60	-.51252	.00005	.00059	.00244	-.00757	-.00507	.00848
8	52.65	-1.39629	.00044	.00069	-.00962	-.02039	.01471	.02727
9	48.48	-.09011	.00000	.00036	-.00012	-.00130	.00092	.00063
10	52.40	-.09098	.00000	.00061	-.00032	.00128	-.00183	.00008
11	47.21	-.93018	.00018	.00064	-.02087	.01351	.01905	-.00301
12	48.39	.34721	.00004	.00096	.00550	-.00511	-.00699	.00636
13	50.19	1.06687	.00039	.00117	.01455	-.01579	-.02136	.02492
14	40.86	1.27964	.00034	.00063	-.02269	.01872	.02621	.00163
15	45.36	.98336	.00013	.00033	.00845	-.01418	.00006	.01105

Visualização de dados Visualização de variável

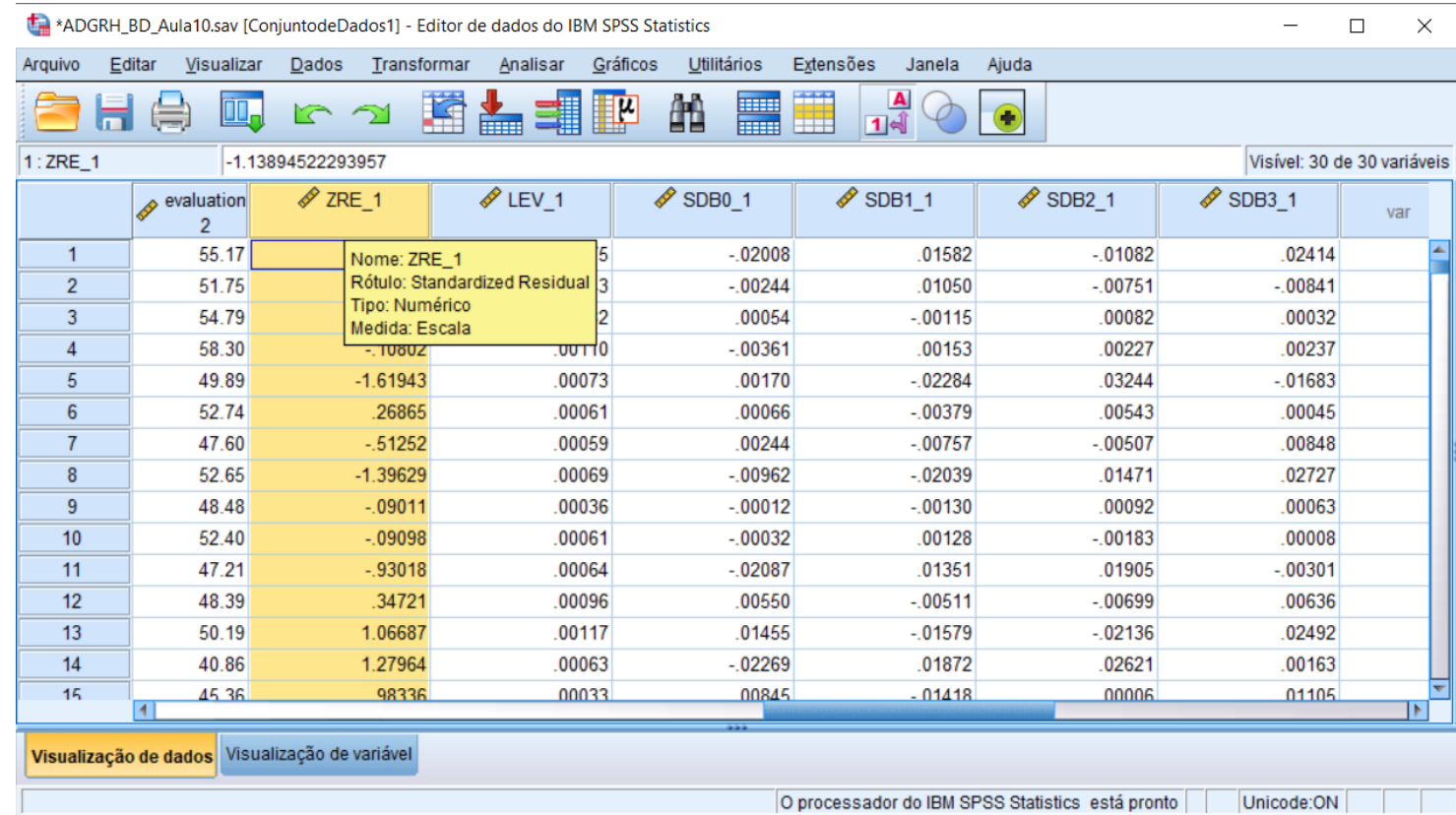
O processador do IBM SPSS Statistics está pronto Unicode:ON

## Validação do Modelo de Regressão Linear

### *2. Avaliação do Pressuposto II: Normalidade da Distribuição dos Erros*

# Normalidade da Distribuição dos Erros

- Para avaliarmos se os erros seguem uma distribuição normal, vamos usar a variável com os 'Resíduos Padronizados' da VD (ZRE\_1) que acabamos de criar.
- Vamos então criar usar um gráfico Q-Q para representar a distribuição dos resíduos padronizados



\*ADGRH\_BD\_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: ZRE\_1 -1.13894522293957 Visível: 30 de 30 variáveis

	evaluation 2	ZRE_1	LEV_1	SDB0_1	SDB1_1	SDB2_1	SDB3_1	var
1	55.17		5	-.02008	.01582	-.01082	.02414	
2	51.75		3	-.00244	.01050	-.00751	-.00841	
3	54.79		2	.00054	-.00115	.00082	.00032	
4	58.30	-.10802	.00110	-.00361	.00153	.00227	.00237	
5	49.89	-1.61943	.00073	.00170	-.02284	.03244	-.01683	
6	52.74	.26865	.00061	.00066	-.00379	.00543	.00045	
7	47.60	-.51252	.00059	.00244	-.00757	-.00507	.00848	
8	52.65	-1.39629	.00069	-.00962	-.02039	.01471	.02727	
9	48.48	-.09011	.00036	-.00012	-.00130	.00092	.00063	
10	52.40	-.09098	.00061	-.00032	.00128	-.00183	.00008	
11	47.21	-.93018	.00064	-.02087	.01351	.01905	-.00301	
12	48.39	.34721	.00096	.00550	-.00511	-.00699	.00636	
13	50.19	1.06687	.00117	.01455	-.01579	-.02136	.02492	
14	40.86	1.27964	.00063	-.02269	.01872	.02621	.00163	
15	45.36	.98336	.00033	.00845	-.01418	.00006	.01105	

Nome: ZRE\_1  
Rótulo: Standardized Residual  
Tipo: Numérico  
Medida: Escala

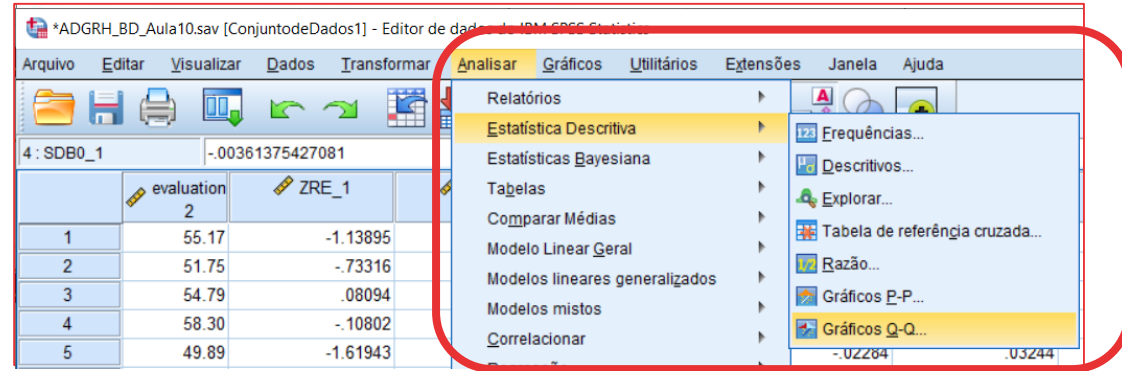
Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

# Normalidade da Distribuição dos Erros

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Gráficos Q-Q'
- Selecionar a variável 'ZRE\_1'
- Colocar na caixa 'Variáveis'
- Selecionar 'OK'

A

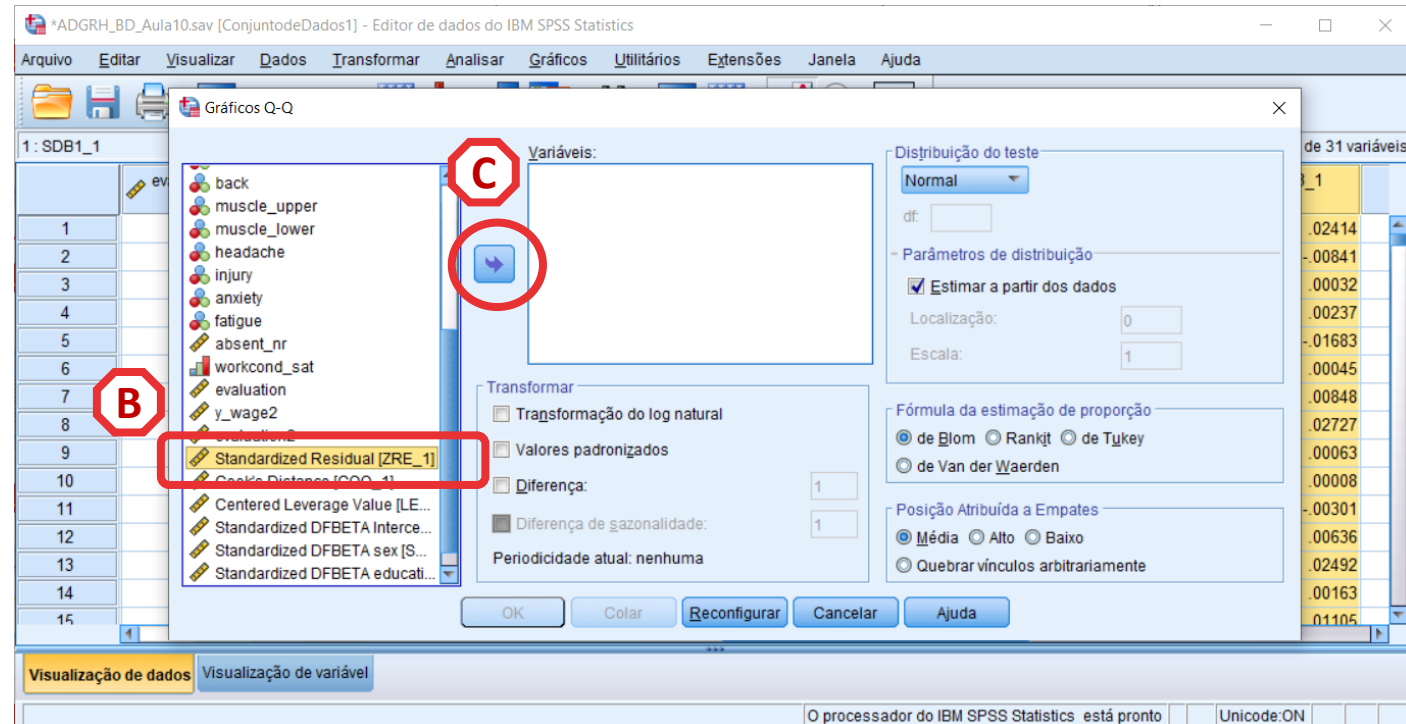


A

B

C

D



# Normalidade da Distribuição dos Erros

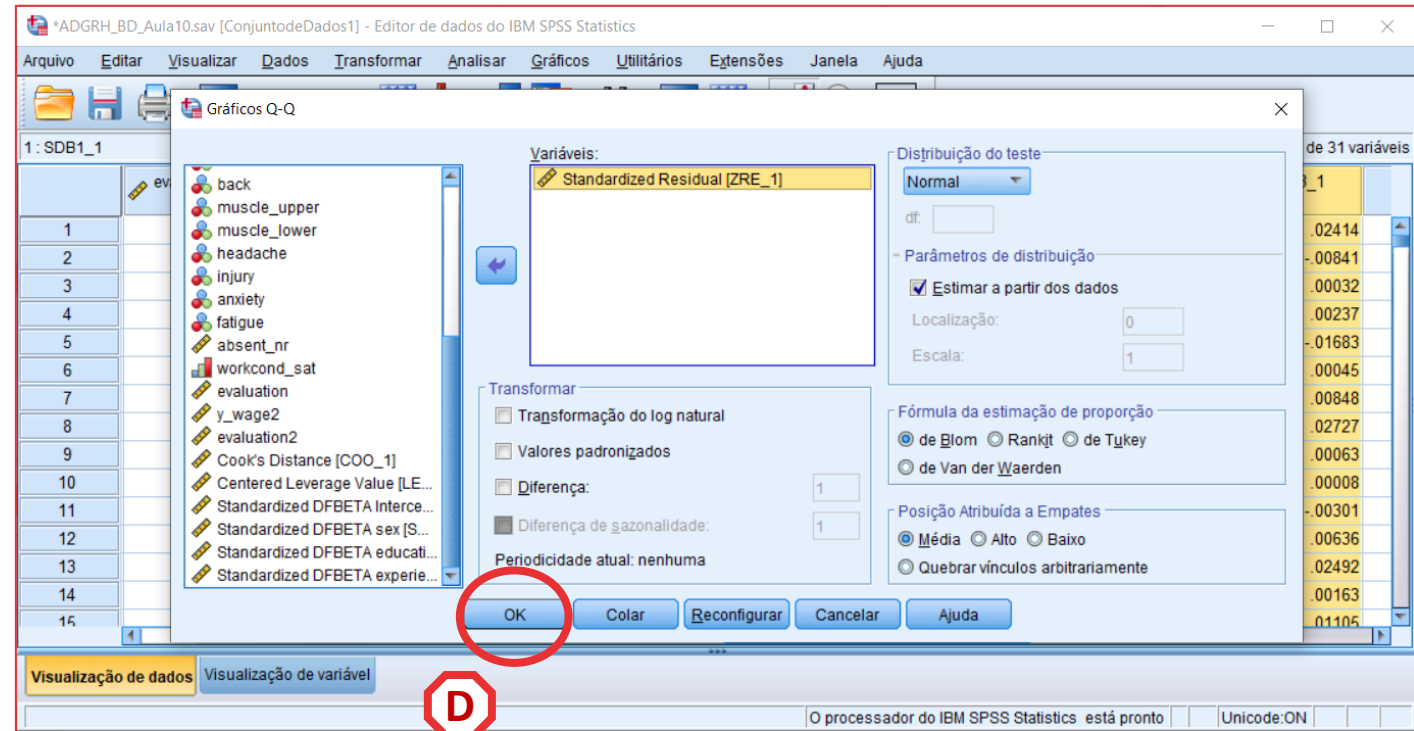
- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Gráficos Q-Q'
- Selecionar a variável 'ZRE\_1'
- Colocar na caixa 'Variáveis'
- Selecionar 'OK'

A

B

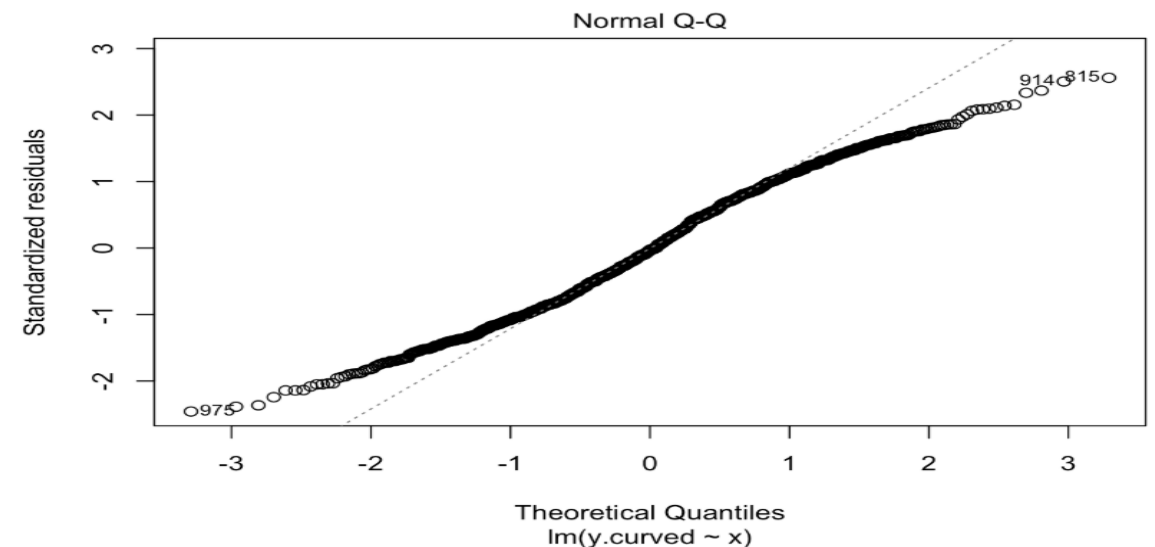
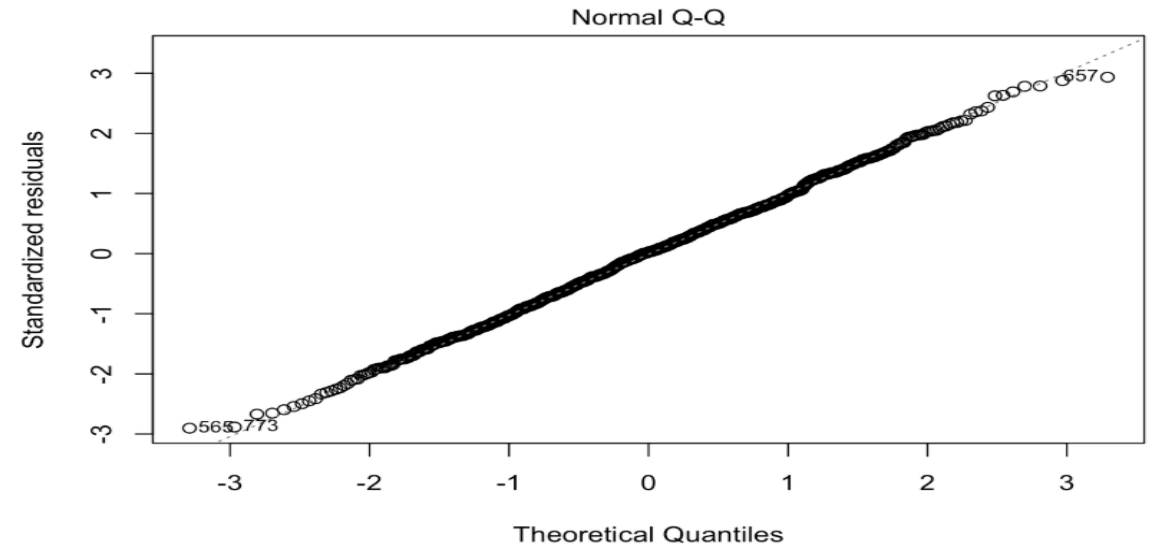
C

D



# Normalidade da Distribuição dos Erros

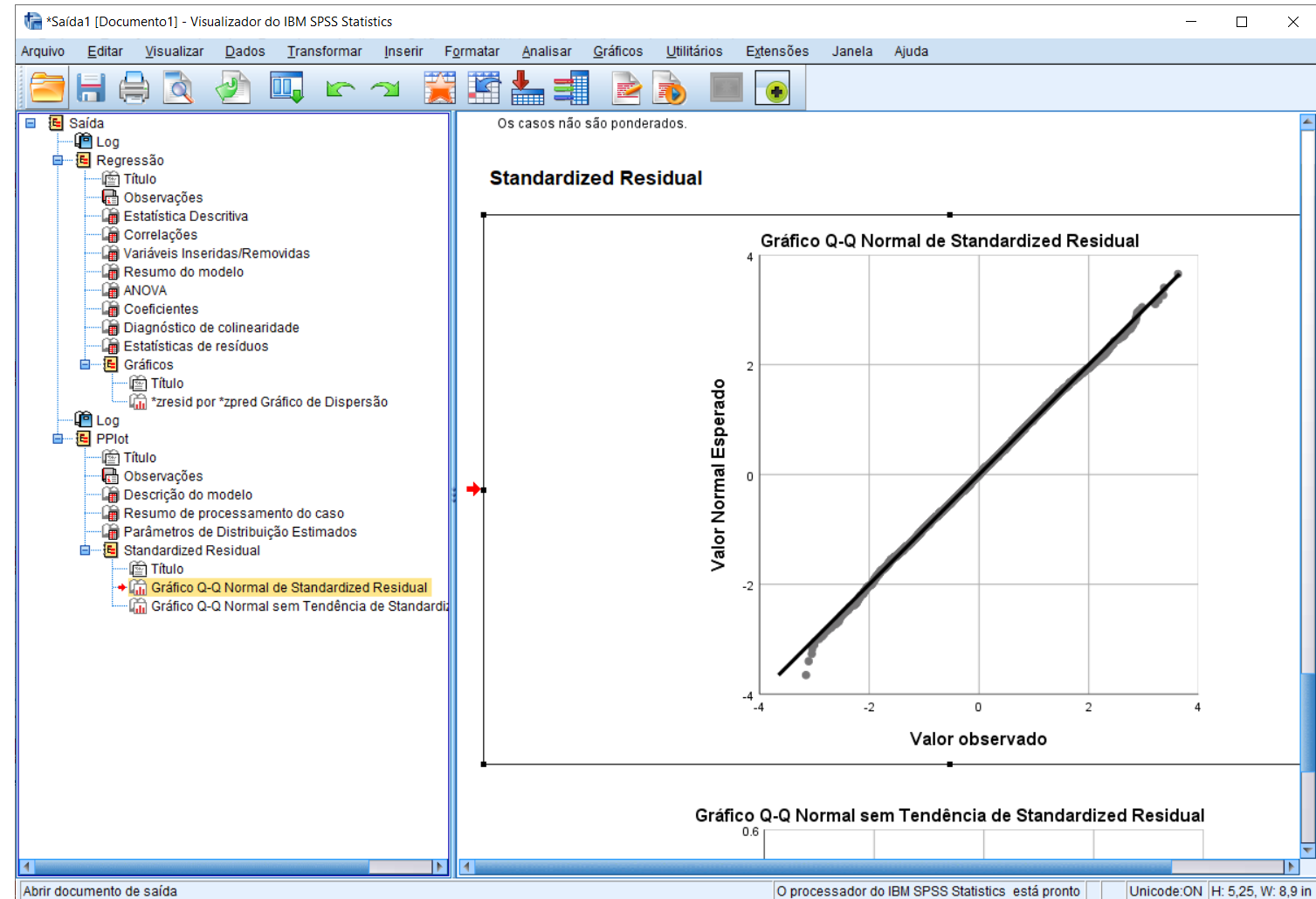
- Linha diagonal reflecte uma distribuição normal
- Os resíduos sobrepõe-se quase totalmente com a linha de diagonal
- Os resíduos parecem estar normalmente distribuídos
- Neste, caso os as caudas da distribuição dos resíduos afasta-se da diagonal, o que sugere que a distribuição dos erros pode não ser normal





# Normalidade da Distribuição dos Erros

- O gráfico é publicado no 'Visualizador de Resultados'
- Neste caso podemos concluir que os erros seguem uma distribuição normal!



# Validação do Modelo de Regressão Linear

## *3. Avaliação do Pressuposto III: Média Condicional Zero dos Erros*

# Média Condicional Zero dos Erros

- Para avaliarmos se o termo de erro aleatório tem valor esperado igual a zero, vamos usar a variável com os 'Resíduos Padronizados' da VD (ZRE\_1) que acabamos de criar.
- Mas neste caso, vamos olhar para as estatísticas descritivas desta variável.

\*ADGRH\_BD\_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: ZRE\_1 -1.13894522293957 Visível: 30 de 30 variáveis

	evaluation 2	ZRE_1	LEV_1	SDB0_1	SDB1_1	SDB2_1	SDB3_1	var
1	55.17		5	-.02008	.01582	-.01082	.02414	
2	51.75		3	-.00244	.01050	-.00751	-.00841	
3	54.79		2	.00054	-.00115	.00082	.00032	
4	58.30	-.10802	.00110	-.00361	.00153	.00227	.00237	
5	49.89	-1.61943	.00073	.00170	-.02284	.03244	-.01683	
6	52.74	.26865	.00061	.00066	-.00379	.00543	.00045	
7	47.60	-.51252	.00059	.00244	-.00757	-.00507	.00848	
8	52.65	-1.39629	.00069	-.00962	-.02039	.01471	.02727	
9	48.48	-.09011	.00036	-.00012	-.00130	.00092	.00063	
10	52.40	-.09098	.00061	-.00032	.00128	-.00183	.00008	
11	47.21	-.93018	.00064	-.02087	.01351	.01905	-.00301	
12	48.39	.34721	.00096	.00550	-.00511	-.00699	.00636	
13	50.19	1.06687	.00117	.01455	-.01579	-.02136	.02492	
14	40.86	1.27964	.00063	-.02269	.01872	.02621	.00163	
15	45.36	.98336	.00033	.00845	-.01418	.00006	.01105	

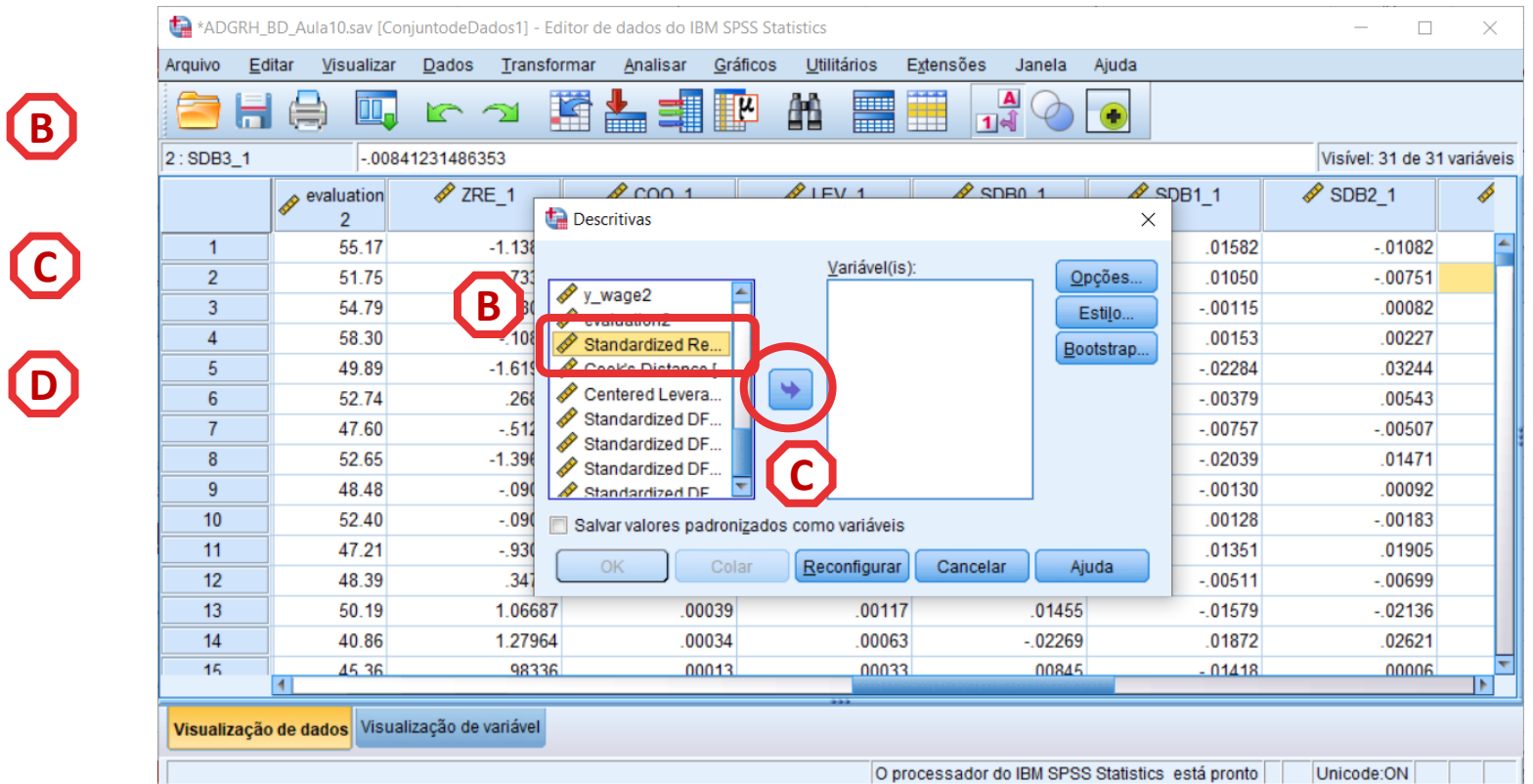
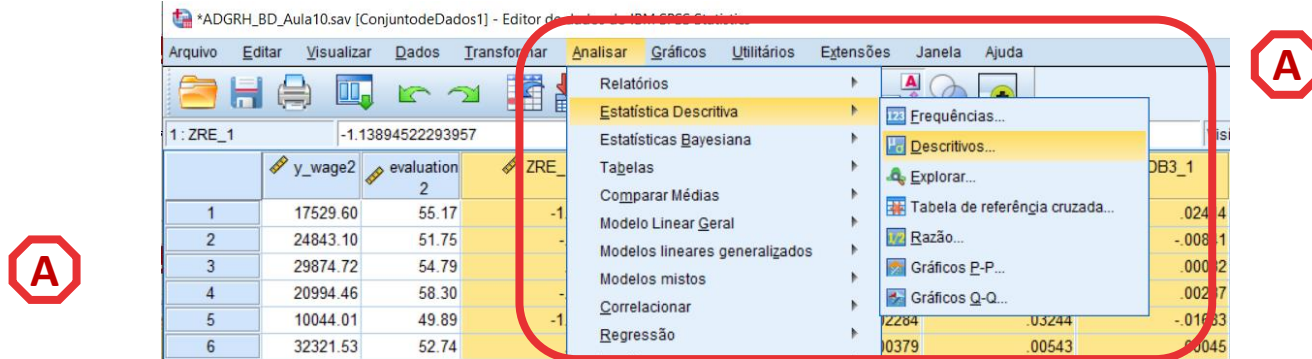
Nome: ZRE\_1  
Rótulo: Standardized Residual  
Tipo: Numérico  
Medida: Escala

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

# Média Condicional Zero dos Erros

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Descritivos'
- Selecionar a variável 'ZRE\_1'
- Colocar na caixa 'Variável(is)'
- Selecionar 'OK'



# Média Condicional Zero dos Erros

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Descritivos'
- Selecionar a variável 'ZRE\_1'
- Colocar na caixa 'Variável(is)'
- Selecionar 'OK'



\*ADGRH\_BD\_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

2: SDB3\_1 -0.00841231486353 Visível: 31 de 31 variáveis

	evaluation 2	ZRE_1	COO_1	LEV_1	SDB0_1	SDB1_1	SDB2_1
1	55.17	-1.138				.01582	-.01082
2	51.75	-.733				.01050	-.00751
3	54.79	.080				-.00115	.00082
4	58.30	-.108				.00153	.00227
5	49.89	-1.619				-.02284	.03244
6	52.74	.268				-.00379	.00543
7	47.60	-.512				-.00757	-.00507
8	52.65	-1.396				-.02039	.01471
9	48.48	-.090				-.00130	.00092
10	52.40					.00128	-.00183
11	47.21					.01351	.01905
12	48.39	.347				-.00511	-.00699
13	50.19	1.06687	.00039	.00117	.01455	-.01579	-.02136
14	40.86	1.27964	.00034	.00063	-.02269	.01872	.02621
15	45.36	.98336	.00013	.00033	.00845	-.01418	.00006

Descriptivas

Variável(is): Standardized Resid...

OK Colar Reconfigurar Cancelar Ajuda

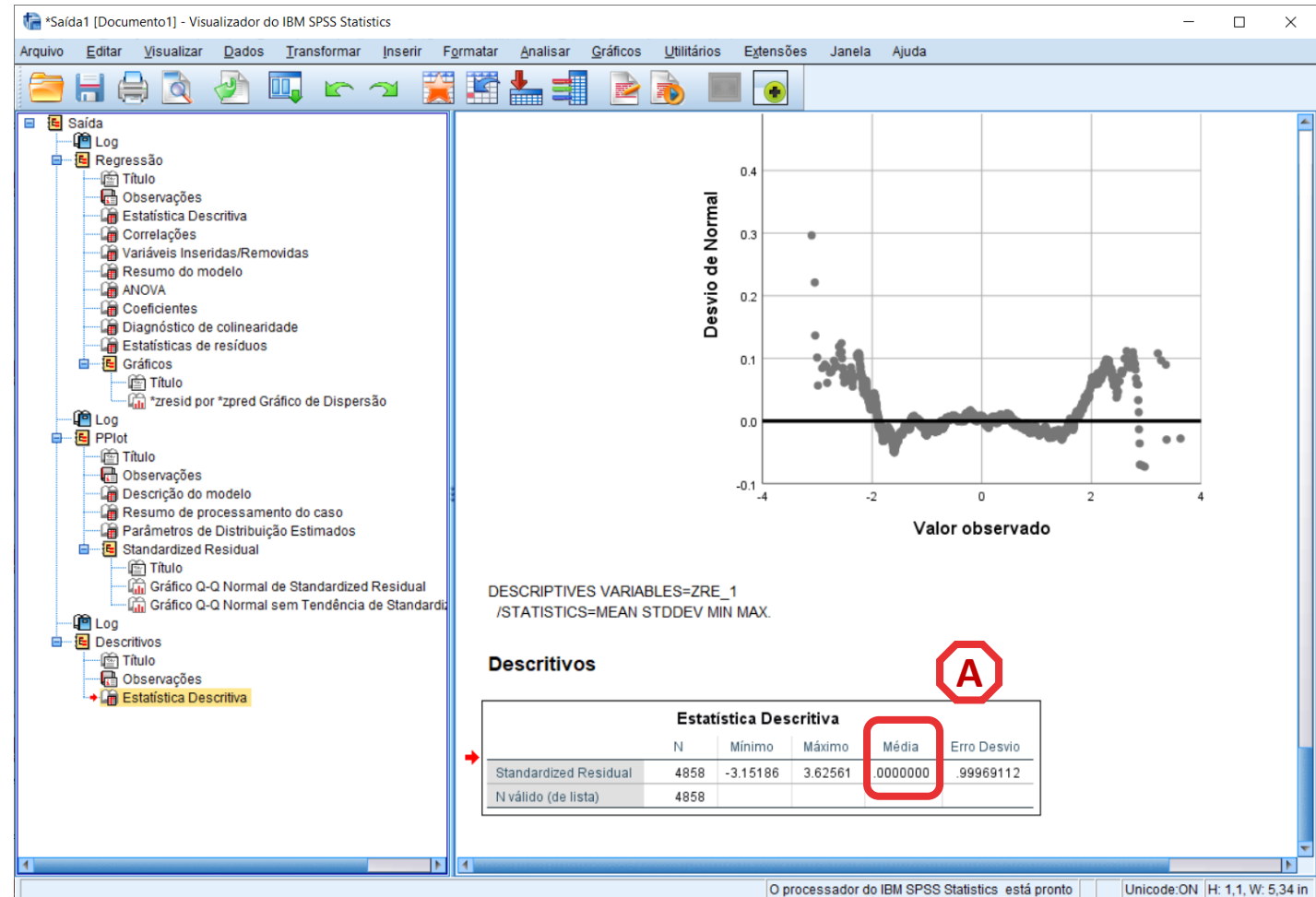
Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

# Média Condicional Zero dos Erros

- O gráfico é publicado no 'Visualizador de Resultados'
- Os 'Resíduos Padronizados' da VD (ZRE\_1) tem uma média muito próximo de 0,
- Neste caso podemos concluir que se cumpre o pressuposto da Média Condicional Zero dos Erros.

A



A

## Validação do Modelo de Regressão Linear

### *4. Avaliação do Pressuposto IV: Homocedasticidade*

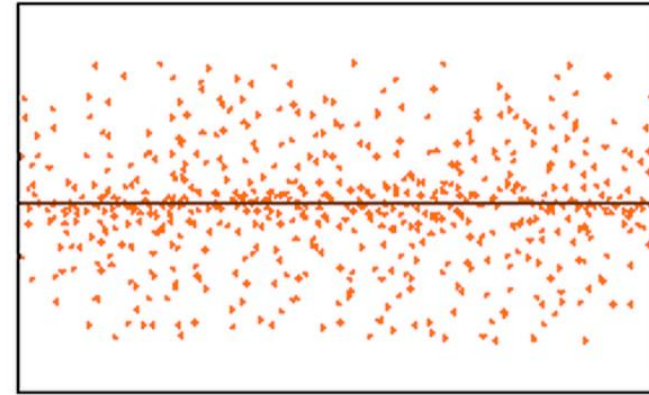
## Homocedasticidade (ou Igual Variância)

- A distribuição dos resíduos apresenta uma variância constante ao longo dos valores previstos da variável dependente. Não há indicação de variação não-constante.

- Neste, o valor dos resíduos aproxima-se de 0 para os valores mais baixos da predição, mas aumentam à medida que os valores previstos também aumenta.

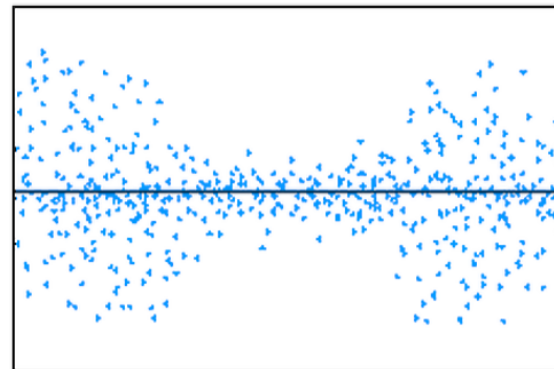
ou seja, a variação não é constante.

Homoscedasticity



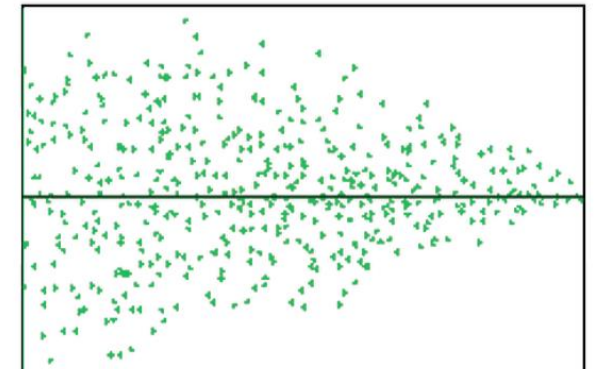
Random Cloud (No Discernible Pattern)

Heteroscedasticity



Bow Tie Shape (Pattern)

Heteroscedasticity



Fan Shape (Pattern)

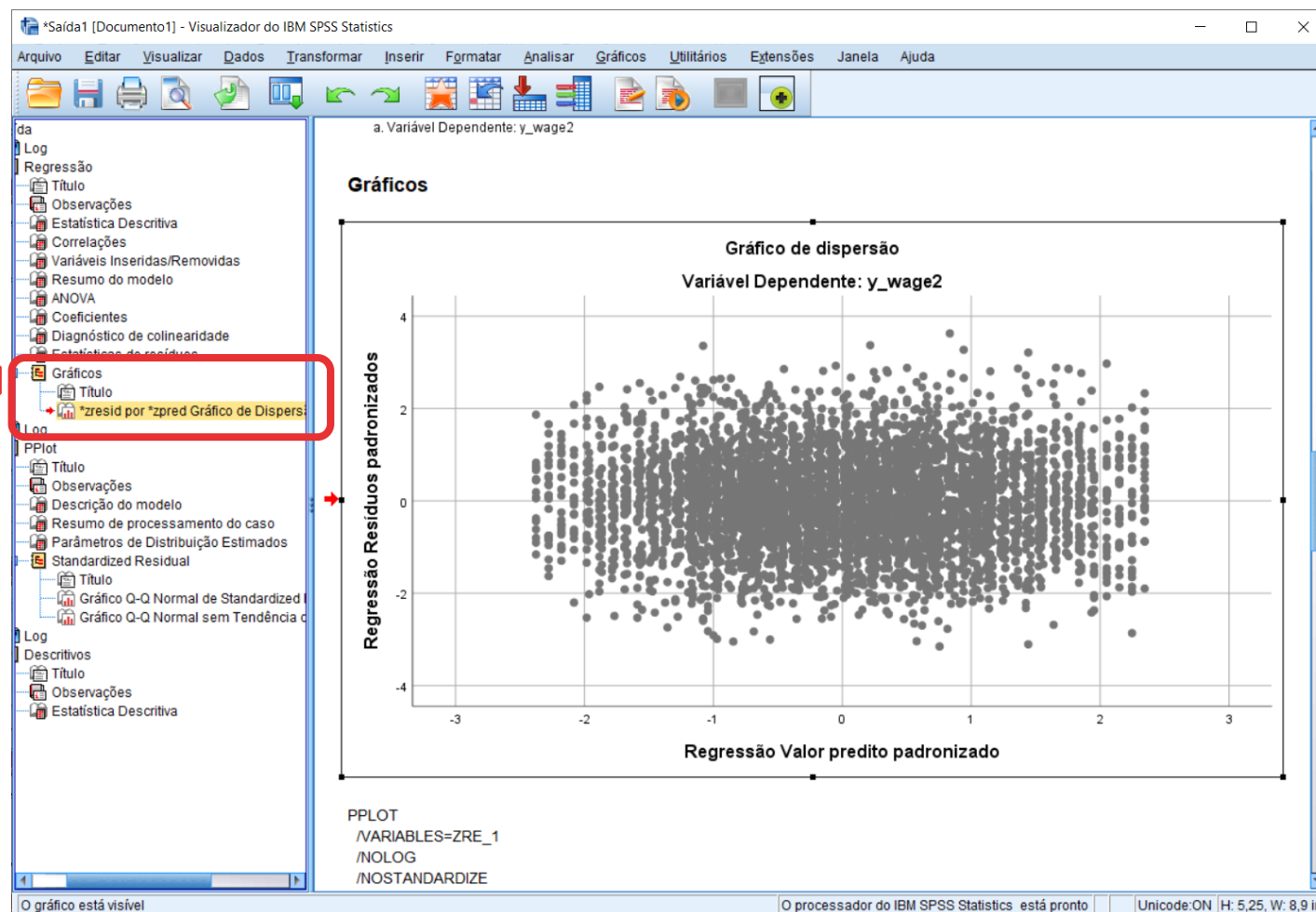


## Homocedasticidade

- Para avaliar se se cumpre este pressuposto, temos de olhar para o Gráfico de Dispersão que compara a distribuição dos 'Resíduos Padronizados' com os 'Valores Preditos Padronizados' - que o SPSS produz automaticamente.
- Neste caso, a representação da distribuição parece sugerir que a variação dos resíduos é relativamente constante.
- Ou seja, cumpre-se o pressuposto da Homocedasticidade

A

A



## Validação do Modelo de Regressão Linear

### *5. Avaliação do Pressuposto V: Independência dos Erros*

# Independência dos Erros

- Para avaliar se se cumpre este pressuposto, temos de olhar para o resultado do teste Durbin-Watson - que pedimos ao SPSS para produzir.

- Interpretação:

= 2 -> Erros são independentes

> 2 / < 2 -> Erros não são independentes

- Neste caso os erros são independentes



**variáveis inseridas/removidas**

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

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

**Resumo do modelo<sup>b</sup>**

Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa	Durbin-Watson
1	.415 <sup>a</sup>	.172	.172	7402.32530	2.024

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

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

Modelo		Soma dos Quadrados	df	Quadrado Médio	Z	Sig.
1	Regressão	5.527E+10	3	1.842E+10	336.203	.000 <sup>b</sup>
	Resíduo	2.660E+11	4854	54794419.89		
	Total	3.212E+11	4857			

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

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

Modelo		Coefficients não padronizados		Coefficients padronizados		t	Sig.	Estatísticas de colinearidade	
		B	Erro Erro	Beta				Tolerância	VIF
1	(Constante)	23814.432	451.902			52.698	.000		
	sex	-4072.228	212.441	-.250		-19.169	.000	1.000	1.000



## Validação do Modelo de Regressão Linear

*6. Avaliação do Pressuposto VI: Ausência de Multicolinearidade Perfeita*

## Ausência de Multicolinearidade Perfeita

- Quando há fortes relações lineares entre os preditores numa regressão, a precisão dos coeficientes de regressão diminui em comparação com o que teria sido se os preditores não se correlacionassem entre si
- Um valor de VIF  $> 3$  sugere a existência de colinearidade no modelo
- Um valor de VIF  $> 10$  sugere a existência de colinearidade séria
  
- Deve repensar-se as variáveis a incluir no modelo

## Ausência de Multicolinearidade

- Para testarmos este pressuposto, temos de olhar para a Tabela de Coeficientes - que o SPSS produz automaticamente.
  - Interpretação
- VIF > 3 -> presença de colinearidade
- Neste caso, não se identifica a presença de colinearidade...
  - Portanto, cumpre-se o pressuposto da ausência de Multicolinearidade

The screenshot displays the IBM SPSS Statistics interface with the following data:

**Modelo**  
Total: 3.212E+11, 4857  
a. Variável Dependente: y\_wage2  
b. Preditores: (Constante), experience, sex, education

Modelo		Coeficientes não padronizados		Coeficientes padronizados		t	Sig.	Estatísticas de colinearidade	
		B	Erro Erro	Beta				Tolerância	VIF
1	(Constante)	23814.432	451.902			52.698	.000		
	sex	-4072.228	212.441	-.250		-19.169	.000	1.000	1.000
	education	1388.801	75.323	.241		18.438	.000	.999	1.001
	experience	331.517	18.584	.233		17.839	.000	.999	1.001

a. Variável Dependente: y\_wage2

Modelo	Dimensão	Autovalor	Índice de condição	Proporções de variância			
				(Constante)	sex	education	experience
1	1	3.612	1.000	.00	.01	.01	.01
	2	.211	4.139	.00	.01	.30	.68
	3	.138	5.115	.02	.35	.48	.18
	4	.039	9.654	.98	.63	.21	.13

a. Variável Dependente: y\_wage2

	Mínimo	Máximo	Média	Erro Desvio	N
Valor previsto	17390.2930	33316.5391	25405.6559	3373.22893	4858
Erro Valor previsto	-2.376	2.345	.000	1.000	4858
Erro padrão do valor previsto	150.030	280.674	209.876	32.697	4858
Valor previsto ajustado	17370.3906	33336.3047	25405.6766	3373.23731	4858

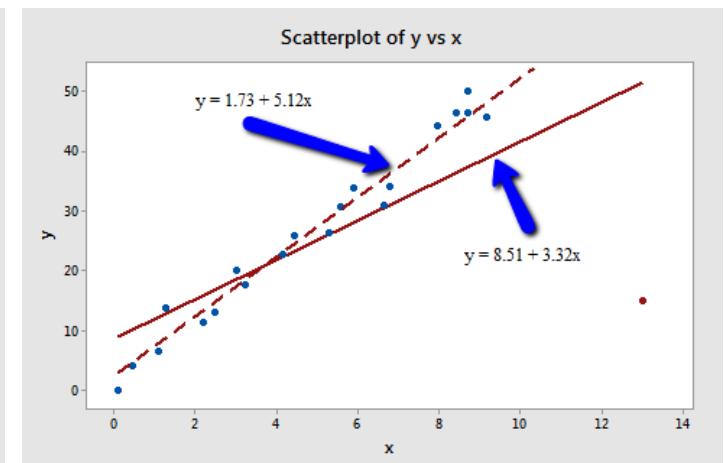
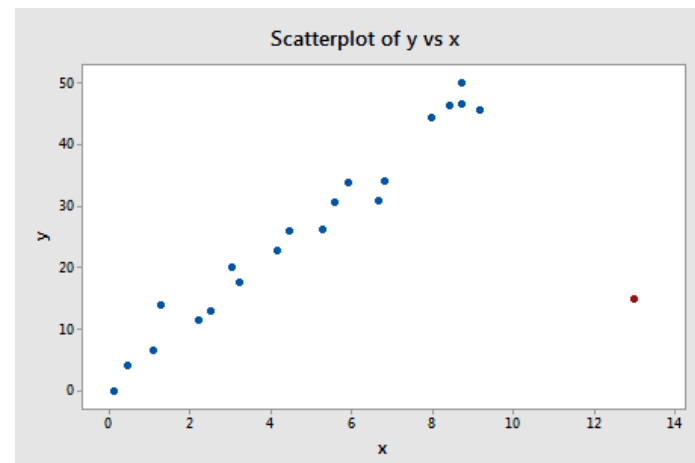
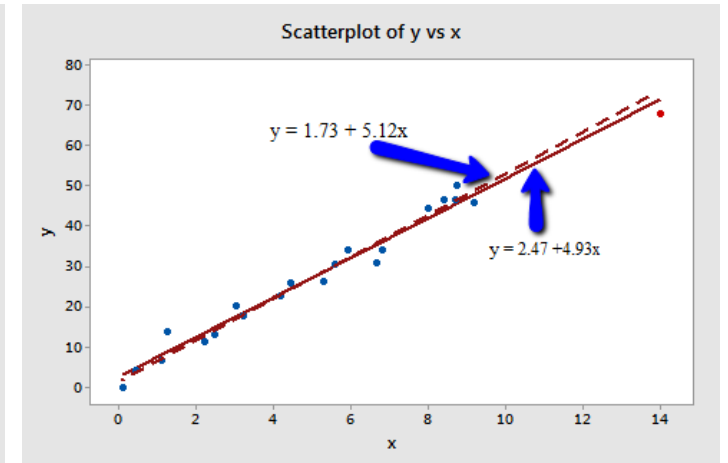
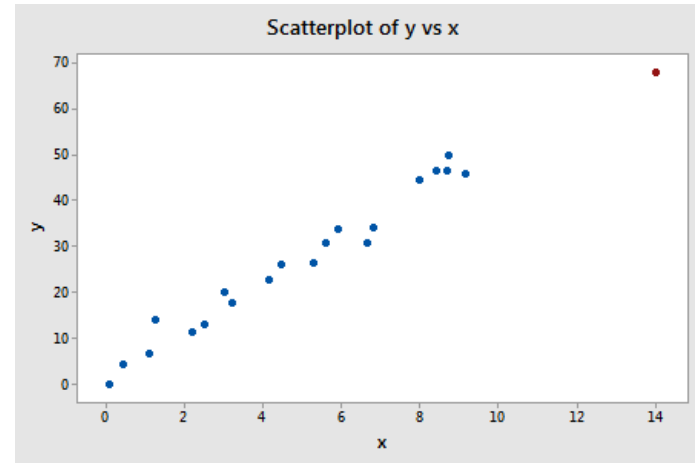


## Validação do Modelo de Regressão Linear

*7. Avaliação do Pressuposto VII: Ausência de Observações Influentes*

## Ausência de Observações Influentes

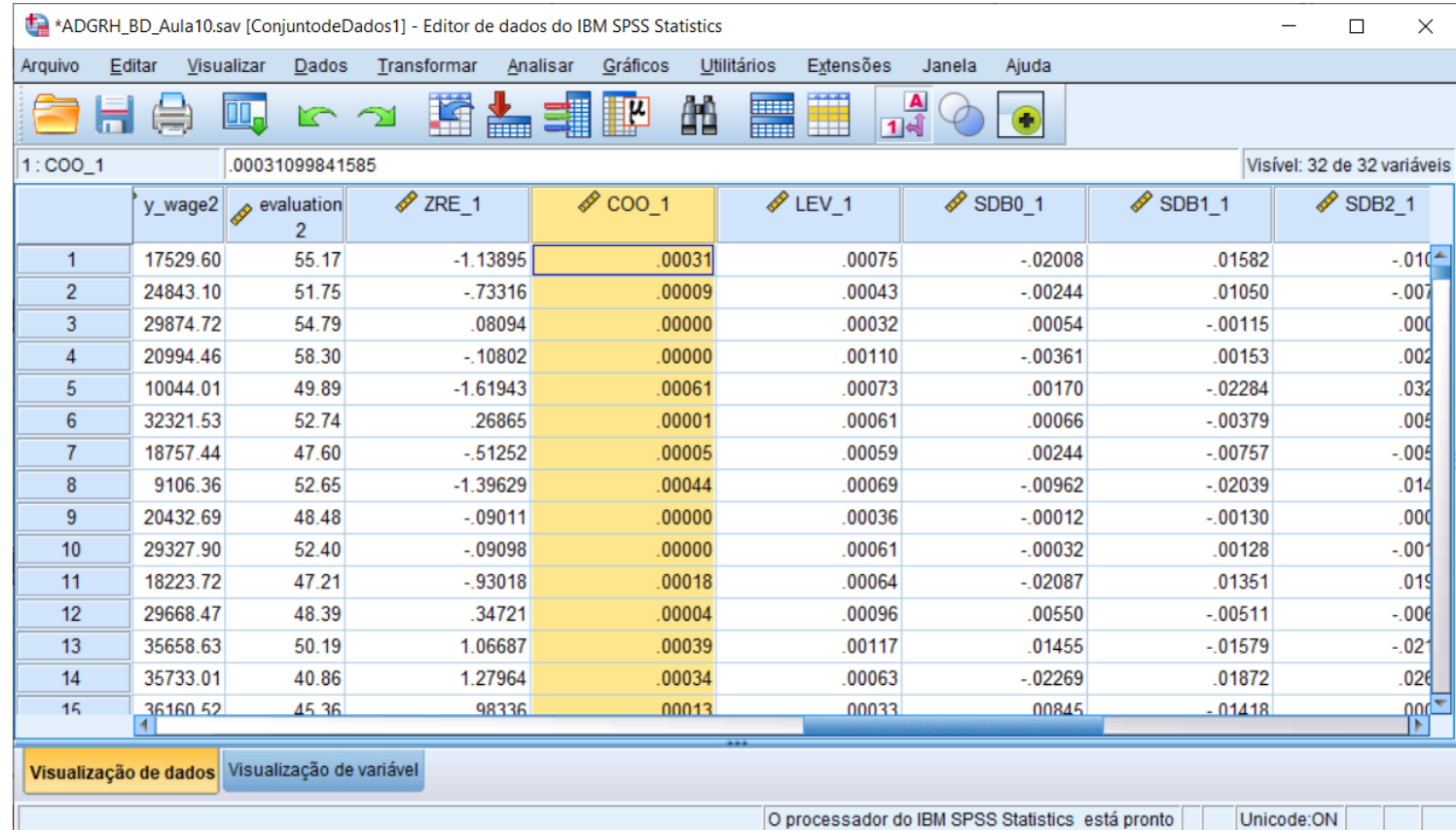
- A existência de 'Outliers' (valores extremos) não é um problema em si.
- Mas torna-se um problema quando os Outlier têm influência sobre os resultados do modelo
- Nos painéis de baixo, o Outlier é uma 'Observação Influyente'





# Ausência de Observações Influentes

- Para testarmos a presença de observações influentes vamos usar a variável com os 'Distância de Cook' (COO\_1) que acabamos de criar.



\*ADGRH\_BD\_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: COO\_1 .00031099841585 Visível: 32 de 32 variáveis

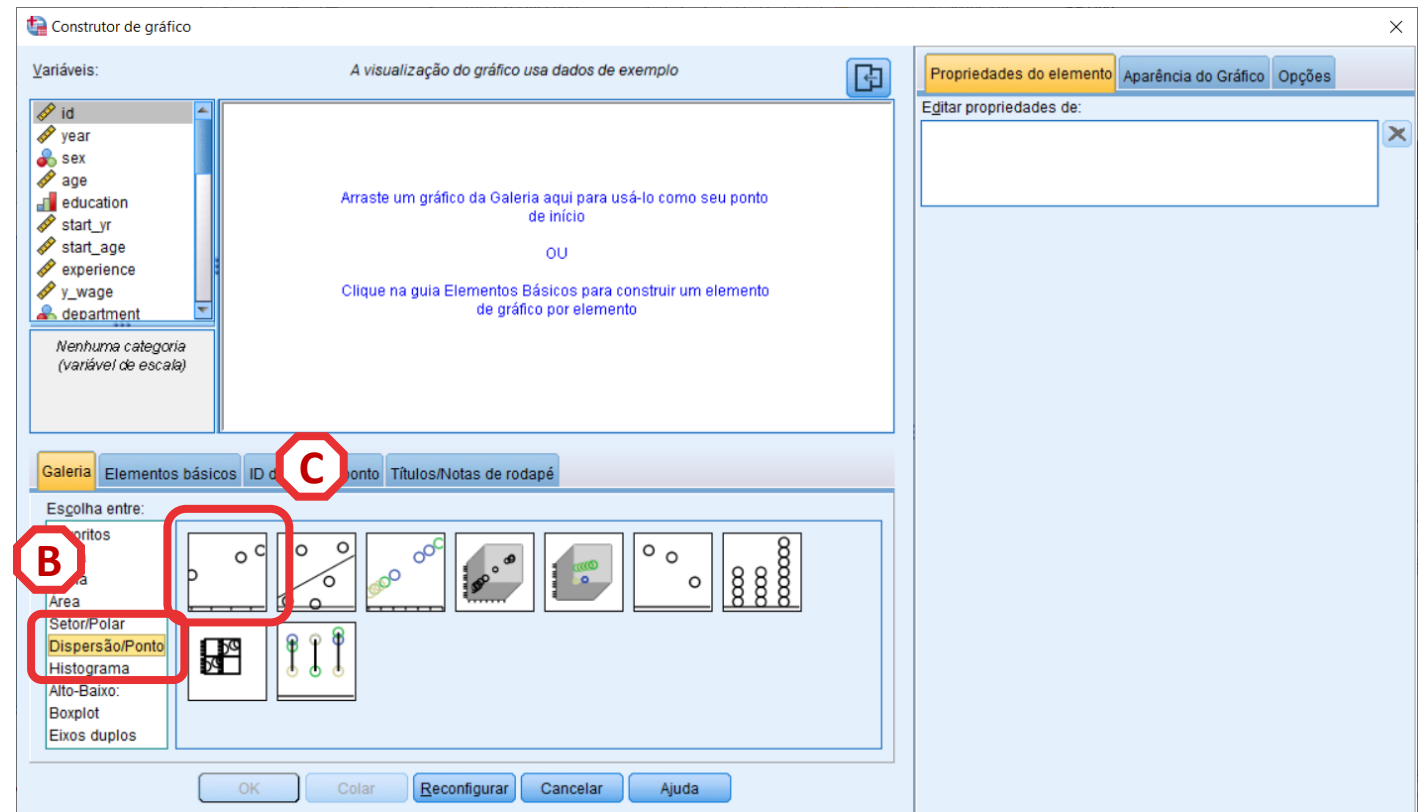
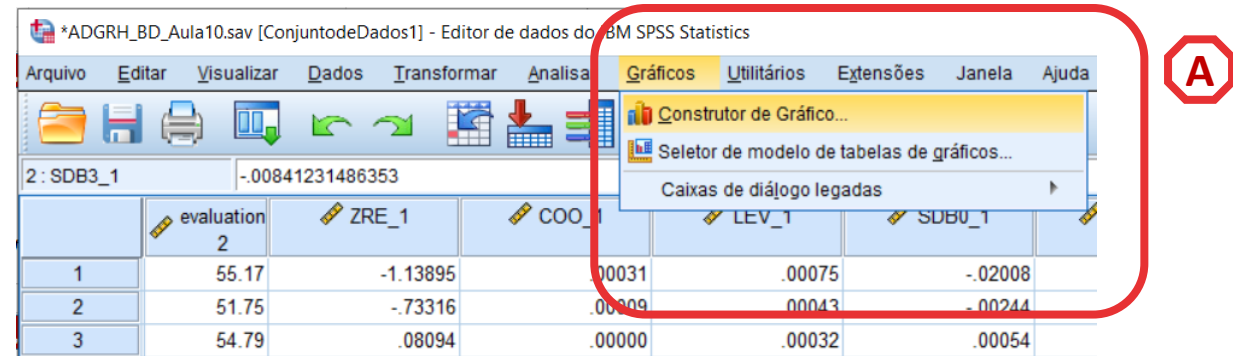
	y_wage2	evaluation 2	ZRE_1	COO_1	LEV_1	SDB0_1	SDB1_1	SDB2_1
1	17529.60	55.17	-1.13895	.00031	.00075	-.02008	.01582	-.010
2	24843.10	51.75	-.73316	.00009	.00043	-.00244	.01050	-.007
3	29874.72	54.79	.08094	.00000	.00032	.00054	-.00115	.000
4	20994.46	58.30	-.10802	.00000	.00110	-.00361	.00153	.002
5	10044.01	49.89	-1.61943	.00061	.00073	.00170	-.02284	.032
6	32321.53	52.74	.26865	.00001	.00061	.00066	-.00379	.005
7	18757.44	47.60	-.51252	.00005	.00059	.00244	-.00757	-.005
8	9106.36	52.65	-1.39629	.00044	.00069	-.00962	-.02039	.014
9	20432.69	48.48	-.09011	.00000	.00036	-.00012	-.00130	.000
10	29327.90	52.40	-.09098	.00000	.00061	-.00032	.00128	-.001
11	18223.72	47.21	-.93018	.00018	.00064	-.02087	.01351	.019
12	29668.47	48.39	.34721	.00004	.00096	.00550	-.00511	-.006
13	35658.63	50.19	1.06687	.00039	.00117	.01455	-.01579	-.021
14	35733.01	40.86	1.27964	.00034	.00063	-.02269	.01872	.026
15	36160.52	45.36	.98336	.00013	.00033	.00845	-.01418	.000

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

# Ausência de Observações Influentes

- Selecionar 'Gráficos' / 'Construtor de Gráfico' A
- Selecionar 'DispersãoPontos' B
- Selecionar 'Dispersão (Simples)' C



# Ausência de Observações Influentes

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'DispersãoPontos'
- Selecionar 'Dispersão (Simples)'
- Selecionar Variável 'id'
- Colocar no eixo 'x'

A

B

C

D

E

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

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

Nenhuma categoria (variável de escala)

Gráfico Disperso 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:

Ponto1

X-Eixo1 (Ponto1)

Y-Eixo1 (Ponto1)

Título 1

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

Empilhar valores idênticos

Exibir linhas de projeção verticais entre pontos

Linhas de Ajuste Lineares

- Total
- Subgrupos

OK | Colar | Reconfigurar | Cancelar | Ajuda

# Ausência de Observações Influentes

- Selecionar 'Gráficos' / 'Construtor de Gráfico' **A**
- Selecionar 'DispersãoPontos' **B**
- Selecionar 'Dispersão (Simples)' **C**
- Selecionar Variável 'id' **D**
- Colocar no eixo 'x' **E**
- Selecionar Variável 'Cooks Distance' **F**
- Colocar no eixo 'y' **G**

The screenshot shows the 'Construtor de gráfico' (Chart Builder) dialog box in SPSS. The window title is 'Construtor de gráfico'. The main area shows a scatter plot titled 'Gráfico Disperso Simples de id'. The X-axis is labeled 'id' and the Y-axis is labeled 'Eixo Y?'. The 'Variáveis:' list on the left includes 'evaluation', 'y\_wage2', 'evaluation2', 'Standardized Re...', 'Cook's Distance [...]', 'Centered Levera...', 'Standardized DF...', 'Standardized DF...', and 'Standardized DF...'. The 'Galeria' tab is selected, showing various chart types. The 'Estatísticas' section on the right is expanded, showing 'Variável:' and 'Estatística:' dropdowns, and a 'Valor' input field. The 'Representação de Barras de Erros' section is also expanded, showing options for 'Intervalos de confiança', 'Erro padrão', and 'Desvio padrão'. The 'Linhas de Ajuste Lineares' section is expanded, showing options for 'Total' and 'Subgrupos'. The 'OK' button is highlighted.

# Ausência de Observações Influentes

- Selecionar 'ID de grupos/ponto'
- Selecionar 'Rótulo da ID do Ponto'
- Selecionar Variável 'id'
- Colocar na caixa 'Variável do rótulo do ponto'
- Selecionar 'OK'



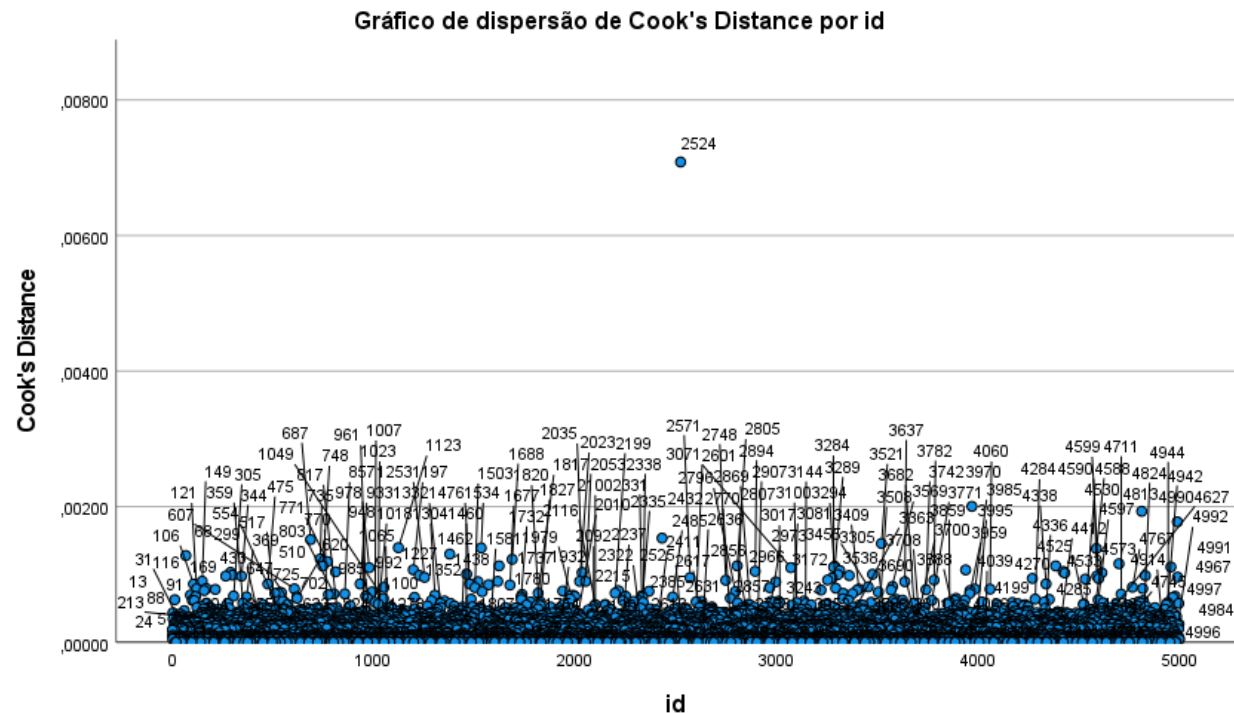
# Ausência de Observações Influentes

- Reparem que o gráfico permite identificar o ID dos outliers

- Interpretação

$CD > 4/n \rightarrow$  Caso Influyente

- Neste caso, o valor de corte é **0.008 (4 / 5000)**
- Neste caso não há observações acima do valor de corte.
- Cumpre-se o pressuposto da ausência de observações influentes



## Reportar os resultados do estudo dos pressupostos

Foram realizados uma série de análises para averiguar a adequabilidade do modelo de regressão linear para o estudo destas relações, sendo que todos os pressupostos assumidos com a aplicação deste técnica foram validados. Em primeiro lugar, analisou-se graficamente a linearidade das relações entre as variáveis independentes (experiência e desempenho) com a variável dependente, tendo sido possível observar relações tendencialmente lineares, especialmente entre desempenho e rendimento (Figura x). Apurou-se também a ausência de multicolinearidade entre as variáveis independentes e de controlo com recurso às medidas VIF ( $<3$ ). Posteriormente, analisou-se a distribuição dos resíduos do modelo, observando-se uma distribuição normal, com um média em torno do valor zero, e com uma variância relativamente constante ao longo dos valores previstos do modelo (Figura x). Os resultados do teste Durbin-Watson, sugerem a ausência de autocorreção significativa nos resíduos ( $D-W=2$ ). Apurou-se ainda a existência de observações influentes com a Distância de Cook, admitindo os valores acima de  $0,008 (4/N)$  como indicadores de observações influentes, não tendo sido detetados casos potencialmente problemáticos à estimação do modelo.

## **Ainda com tempo?**

Repetimos o exercício com o caso da aula passada, incluído a educação no modelo.