



Aula 8: As diferenças de produtividade entre departamentos são estatisticamente significativas?

Formulação e Teste de Hipóteses (II)

Docente: Amílcar Moreira

Data & Hora: 17/11/2020, 20:30-22:30

Local: Edifício F2, Sala 111

- **Na Aula Anterior**

- Fizemos um breve introdução à Estatística Inferencial, por oposição à Estatística Descritiva;
- Discutimos as bases da inferência estatística (Teorema do Limite Central);
- Discutimos o Intervalo Confiança como ferramenta de inferência estatística

- **Objetivos da Aula**

- **Parte Teórica**

- Perceber o que são os Testes de Hipóteses e em que medida se distinguem dos Intervalos de Confiança enquanto ferramenta de inferência estatística;
- Saber Distinguir entre Hipótese Nula e Hipótese Alternativa;
- Saber identificar os critérios para a escolha do Teste de Hipótese adequado;
- Saber Distinguir entre Erros de Tipo I e Erros de Tipo II.

- **Parte Prática**

- Saber aplicar Testes de Hipóteses para os casos mais comuns



• Estatística Descritiva

- **Dá-nos as ferramentas para descrever dados de uma (ou mais variáveis) numa amostra**
 - **Medidas de tendência central (médias, modas, etc.)**
 - **Distribuição de frequências (proporções, percentagens, etc.)**
 - **Medidas de dispersão (variância, desvio padrão, etc.)**

• Estatística Inferencial

- **Dá-nos as ferramentas para avaliarmos se a forma como os dados estão distribuídos na amostra pode ser inferida para a população**
 - **Intervalos de Confiança**
 - **Testes de Hipóteses**

- **Teste de Hipóteses**

- **Procedimento para testar uma afirmação sobre uma propriedade da população:**
 - **A distribuição da variável é normal?**
 - **As diferenças (médias, proporções, etc.) entre grupos são estatisticamente significativas?**
 - **A relação entre as variáveis (associação e correlação) é estatisticamente significativa?**



- **Intervalo de Confiança**

- Fornece um conjunto de valores plausíveis da estimativa (ex. média) na população.

- **Teste de Hipóteses**

- Implica a formulação de hipóteses formais
- Força uma tomada de decisão relativa à significância estatística

- **Teste de Hipóteses**

- **Envolve a formulação de duas hipóteses alternativas**

- **Hipótese Nula (H_0)**

- **Determina o valor do parâmetro da população que se pretende testar (ex. média, proporção, etc.)**
- **Exprime-se sobre a forma de uma igualdade (=)**

- **Hipótese Alternativa (H_1)**

- **Determina que o valor do parâmetro é diferente do que o definido pela Hipótese Nula**
- **Consequentemente pode exprimir-se de uma destas formas**
 - **$\neq H_0$ - Parâmetro é diferente do que é definido pela Hipótese Nula**
 - **$> H_0$ - Parâmetro é maior do que é definido pela Hipótese Nula**
 - **$< H_0$ - Parâmetro é menor do que é definido pela Hipótese Nula**

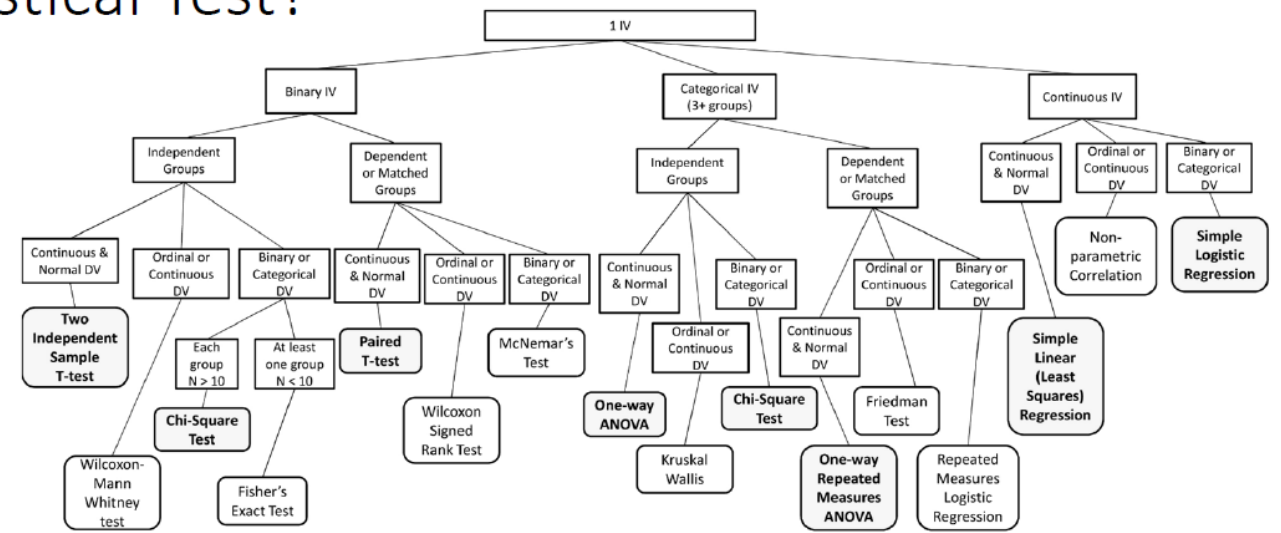
ANÁLISE DE DADOS EM GRH

Aula 8: Formulação e Teste de Hipóteses (II)

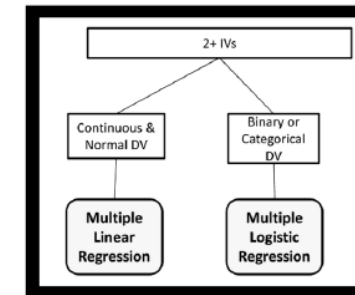
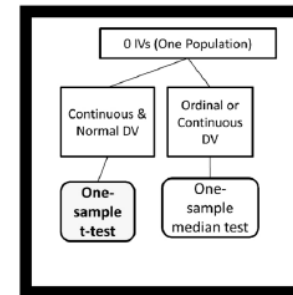
- Como escolher o teste de hipóteses mais adequado?

Which Statistical Test?

1. Number of IVs
2. IV Measurement Scale
3. Independent vs. Matched Groups
4. DV Measurement Scale



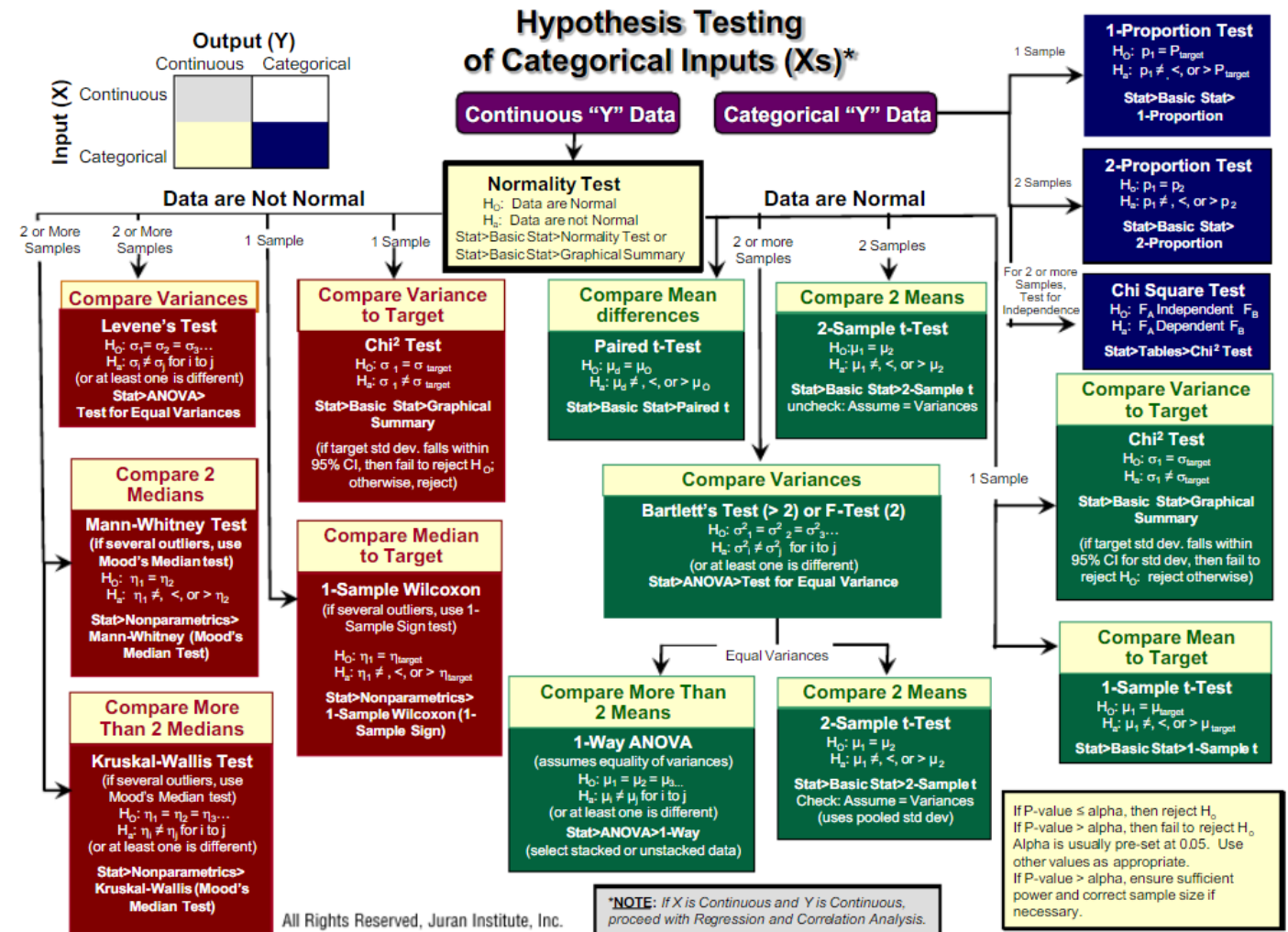
LEGEND:
 IV = Independent Variable (i.e. predictor, exposure)
 DV = Dependent Variable (i.e. response, outcome)



ANÁLISE DE DADOS EM GRH

Aula 8: Formulação e Teste de Hipóteses (II)

- Como escolher o teste de hipóteses mais adequado?



- O que temos de ter em atenção na escolha do Teste de Hipóteses?

Qual é o objectivo?

A estatística amostral (ex: média) é representativa da população?

As diferenças entre grupos na amostra são representativas da população?

1 Amostra

2 Grupos

>2 Grupos

2 Amostras

+2 Amostras

Qual é a escala da variável?

Nominal
(**Proporções**)

Ordinal*

Contínua⁹
(**Médias**)

- O que temos de ter em atenção na escolha do Teste de Significância?

Qual é o tipo de amostra?

Independente

Emparelhada

A amostra segue uma distribuição normal?

Sim
(Teste Paramétrico)

Não
(Teste Não-Paramétrico)

O que diz a Hipótese Alternativa?

$H_1 \neq H_0$
(Teste Bilateral)

$H_1 < H_0$
(Teste Unilateral à Esquerda)

$H_1 > H_0$
(Teste Unilateral à Dtª)

Testes de Hipóteses

A variável segue uma distribuição normal? (Shapiro-Wilk)

A diferença entre médias (2 grupos) é significativa? (Teste de T)

*A diferença entre médias (+2 grupos) é significativa? (ANOVA)**

Há uma relação sistemática entre as variáveis? (Qui-Quadrado)

A diferença entre proporções é significativa? (Teste de Z)

Testes de Hipóteses

A variável segue uma distribuição normal? (Shapiro-Wilk)

Teste de Shapiro-Wilk

- **Objectivo:**
 - **Determinar se a distribuição dos salários na organização ('y_wage2') segue uma distribuição normal?**

Hipótese Nula (H_0):

“A distribuição dos salários na organização segue uma distribuição normal”

Hipótese Alternativa (H_1):

“A distribuição dos salários na organização não segue uma distribuição normal”

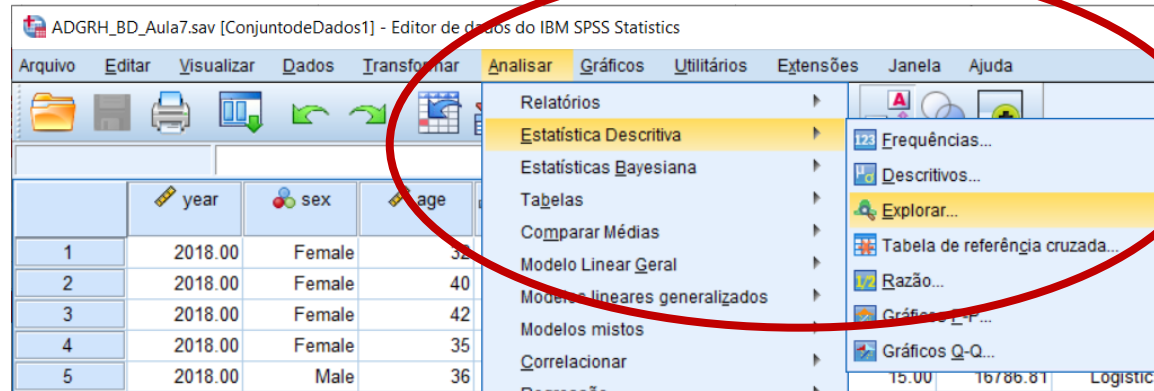
Teste de Shapiro-Wilk

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

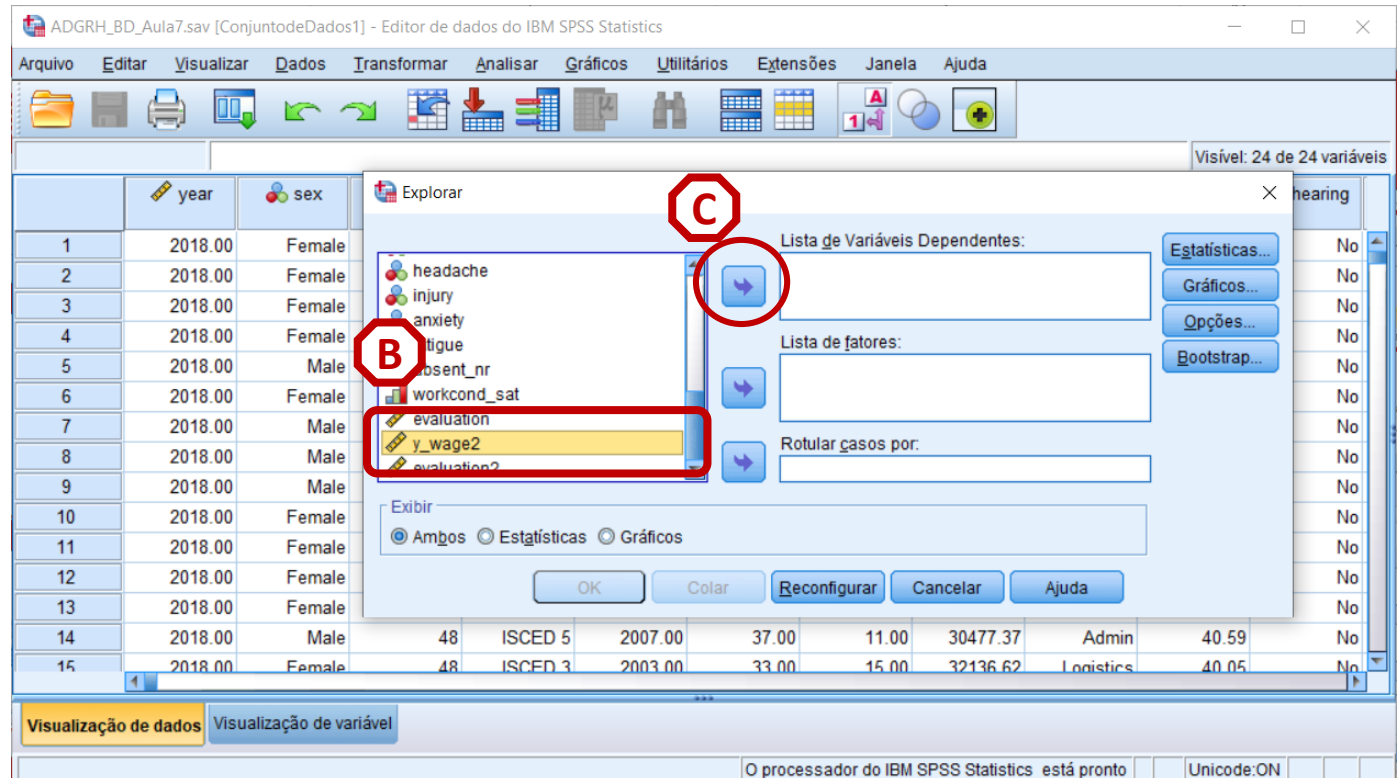
A

B

C



A



Teste de Shapiro-Wilk

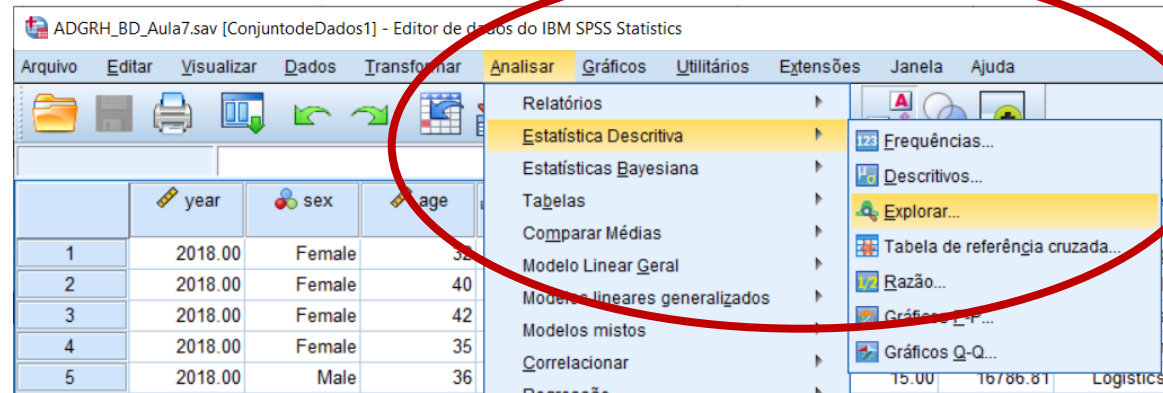
- Selecionar 'Analisar' / 'Estatísticas Descritivas' / 'Explorar'
- Selecionar a variável 'y_wage2'
- Colocar na caixa 'Lista de Variáveis Dependentes'
- Selecionar 'Gráficos'

A

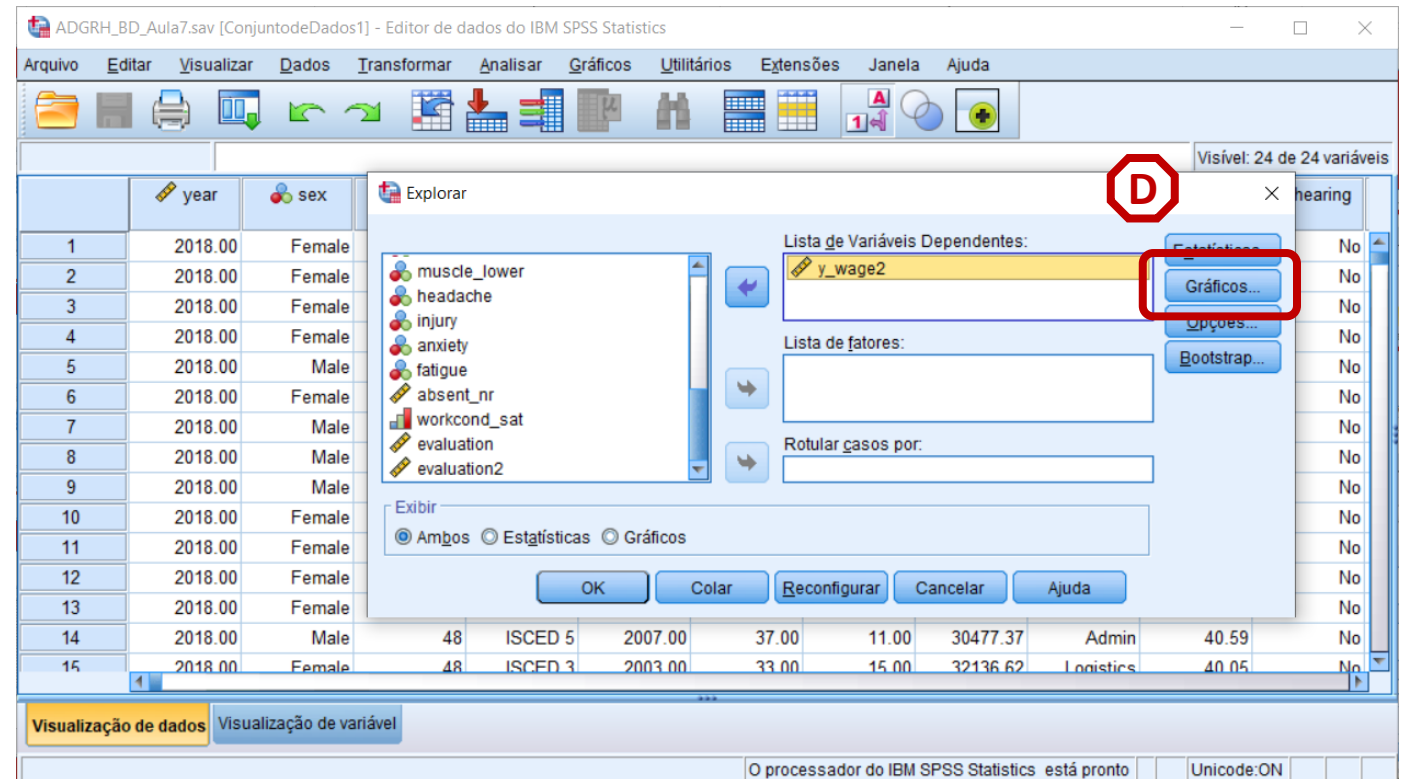
B

C

D



A



D

Teste de Shapiro-Wilk

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

A

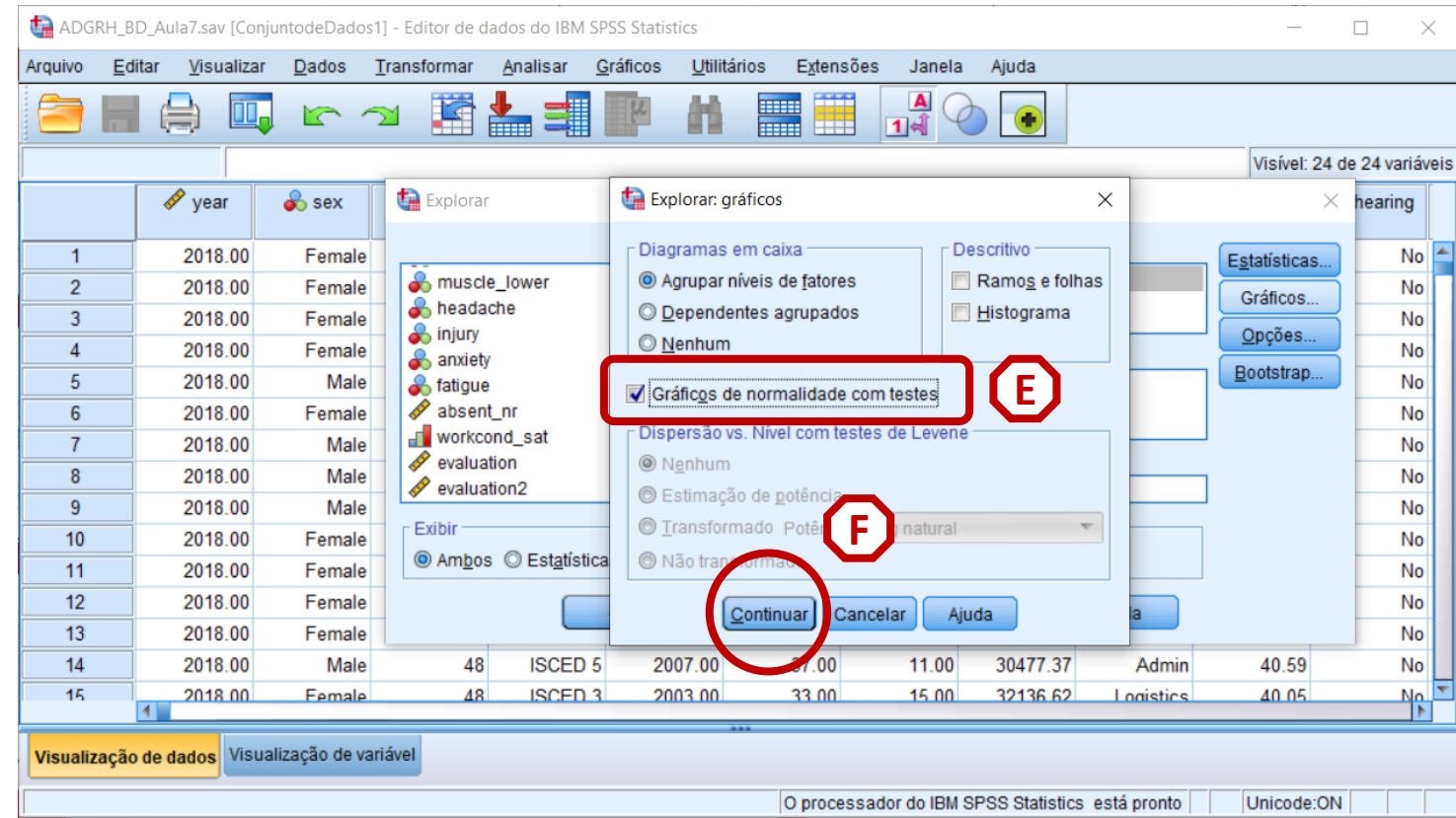
B

C

D

E

F



Teste de Shapiro-Wilk

- Selecionar 'Opções'



ADGRH_BD_Aula7.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Visual: 24 de 24 variáveis

Explorar

Lista de Variáveis Dependentes: y_wage2

Lista de fatores:

Rotular casos por:

Exibir: Ambos Estatísticas Gráficos

OK Colar Reconfigurar Cancelar Ajuda

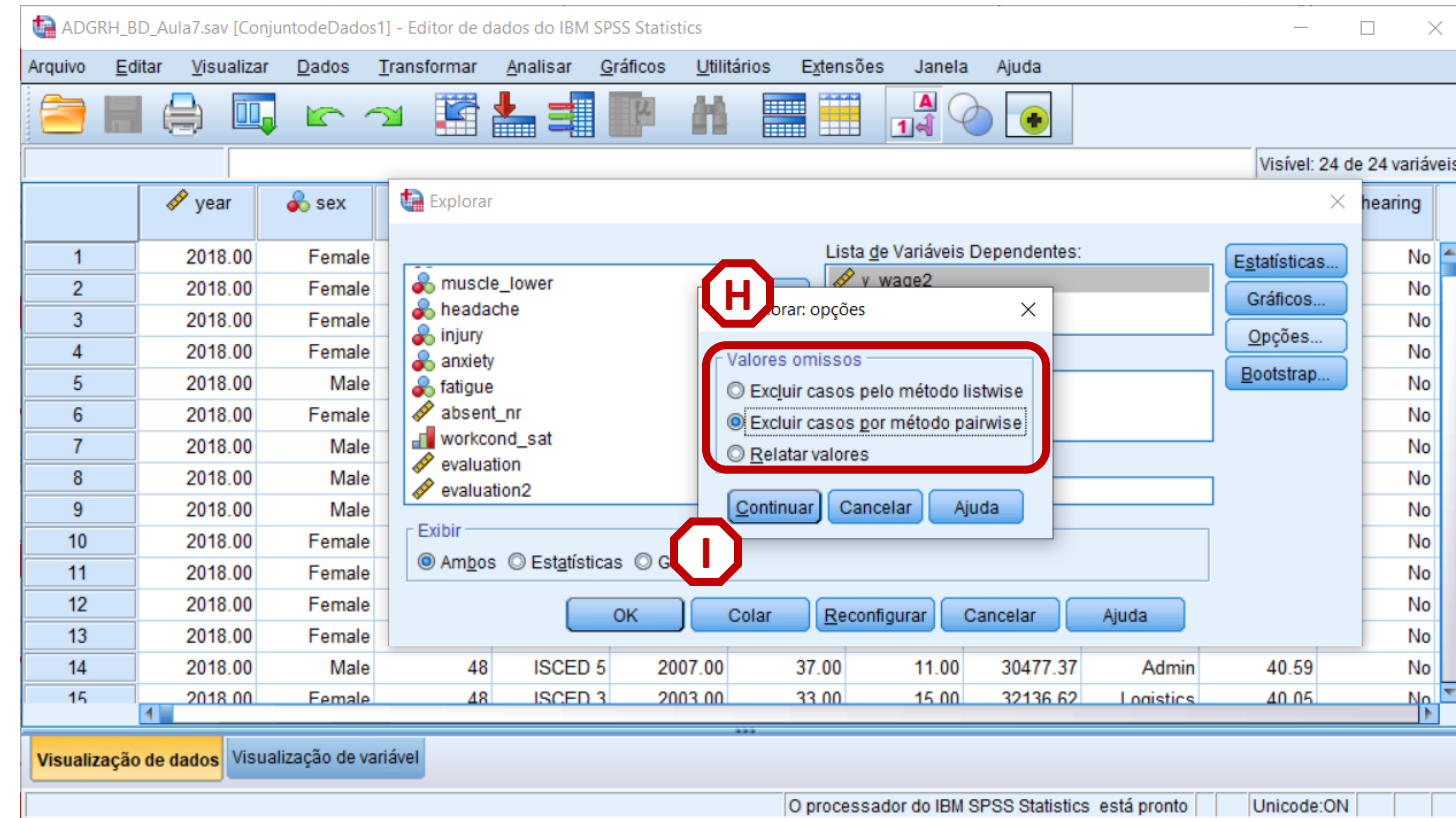
	year	sex	muscle_lower	headache	injury	anxiety	fatigue	absent_nr	workcond_sat	evaluation	evaluation2	hearing
1	2018.00	Female										No
2	2018.00	Female										No
3	2018.00	Female										No
4	2018.00	Female										No
5	2018.00	Male										No
6	2018.00	Female										No
7	2018.00	Male						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

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

Teste de Shapiro-Wilk

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

ADGRH_BD_Aula7.sav [Conjunto de Dados] - 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:

year w wage2

Opções

Valores omissos

Excluir casos pelo método listwise
 Excluir casos por método pairwise
 Relatar valores

Continuar Cancelar Ajuda

Exibir

Ambos Estatísticas Gráficos

OK Colar Reconfigurar Cancelar Ajuda

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

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

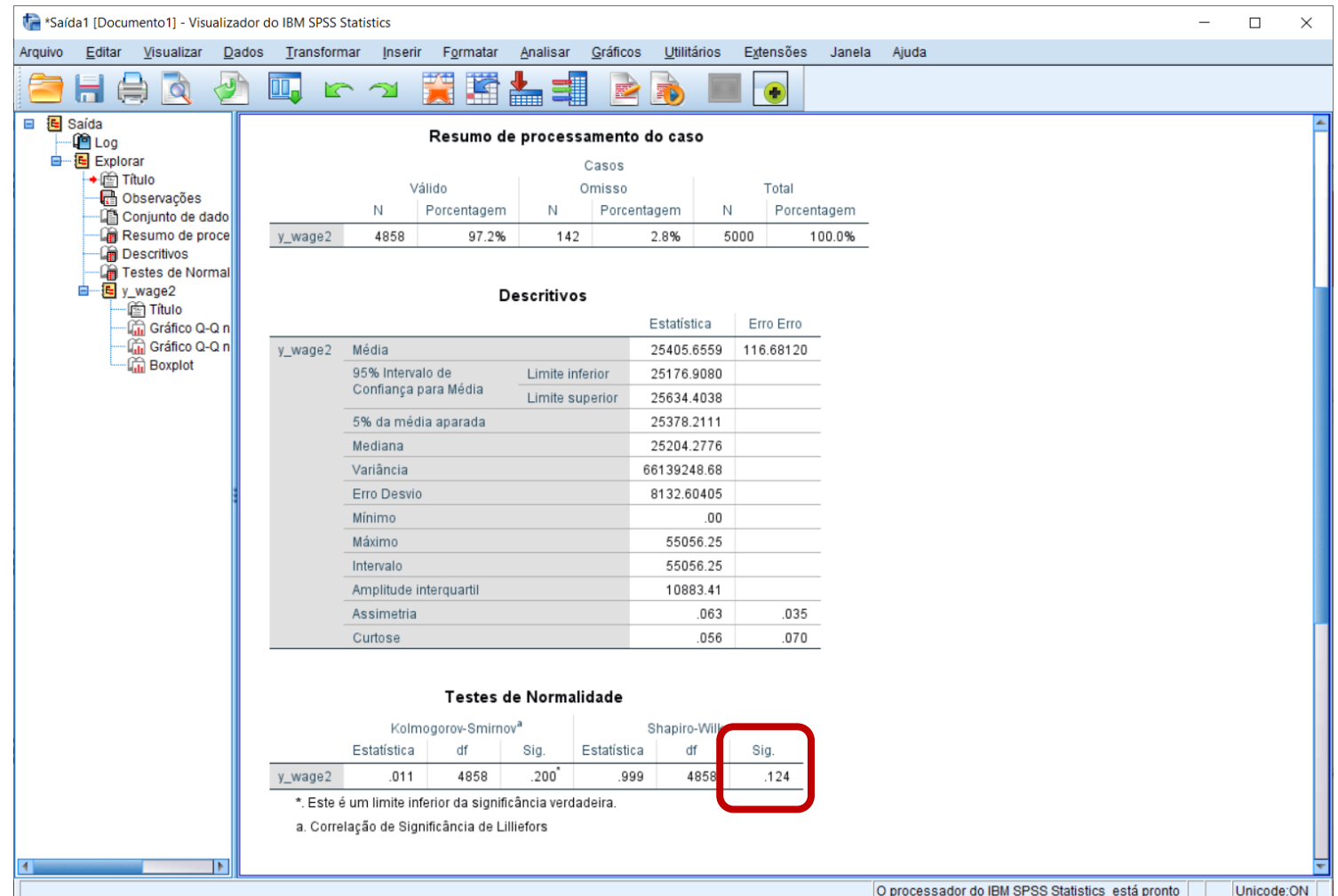
Teste de Shapiro-Wilk

- O resultado é publicado no 'Visualizador de Resultados'

INTERPRETAÇÃO:

- 'Sig'. ≤ 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.



The screenshot shows the IBM SPSS Statistics 'Visualizador de Resultados' window. The main content area displays the following information:

Resumo de processamento do caso

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

Descritivos

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

Testes de Normalidade

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

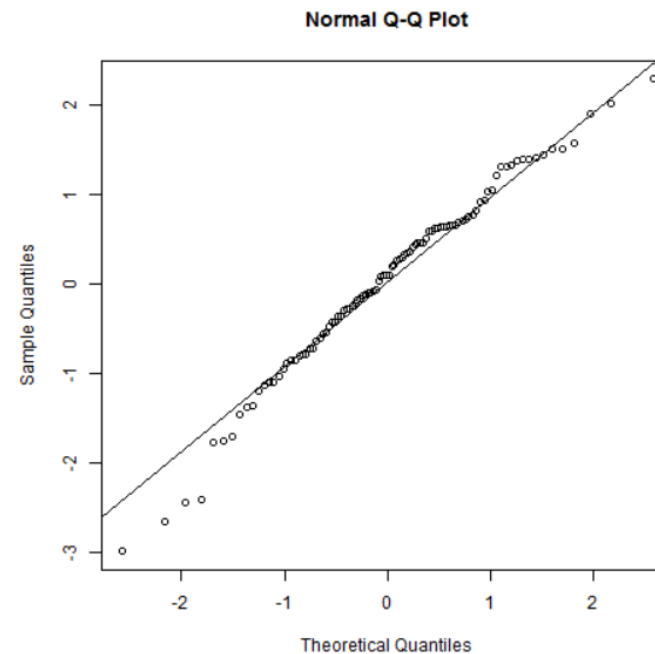
*. Este é um limite inferior da significância verdadeira.
a. Correlação de Significância de Lilliefors

Teste de Shapiro-Wilk

- O SPSS também oferece uma forma de visualizar se a distribuição da variável 'Idade' segue uma distribuição normal:

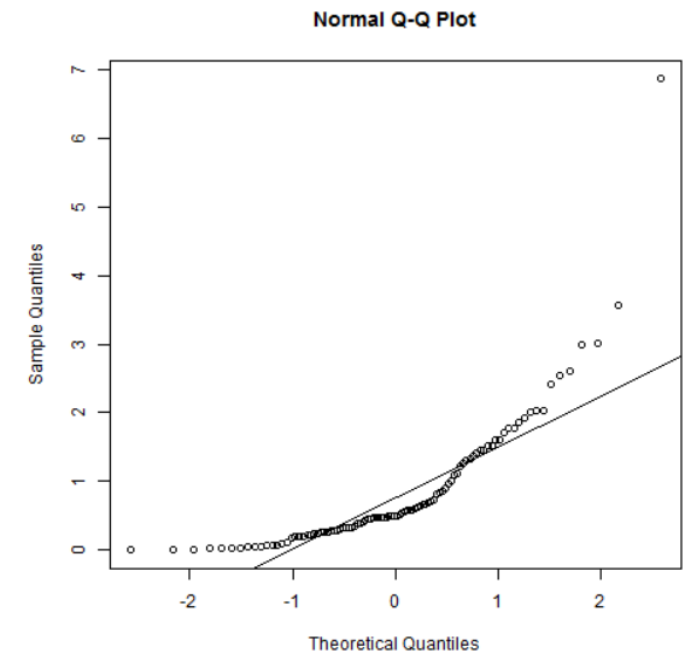
O gráfico Q-Q

Se os pontos se distribuem mais ou menos ao longo da linha...



... Podemos assumir que a variável segue uma distribuição normal.

Se os pontos seguem uma forma distinta da linha...



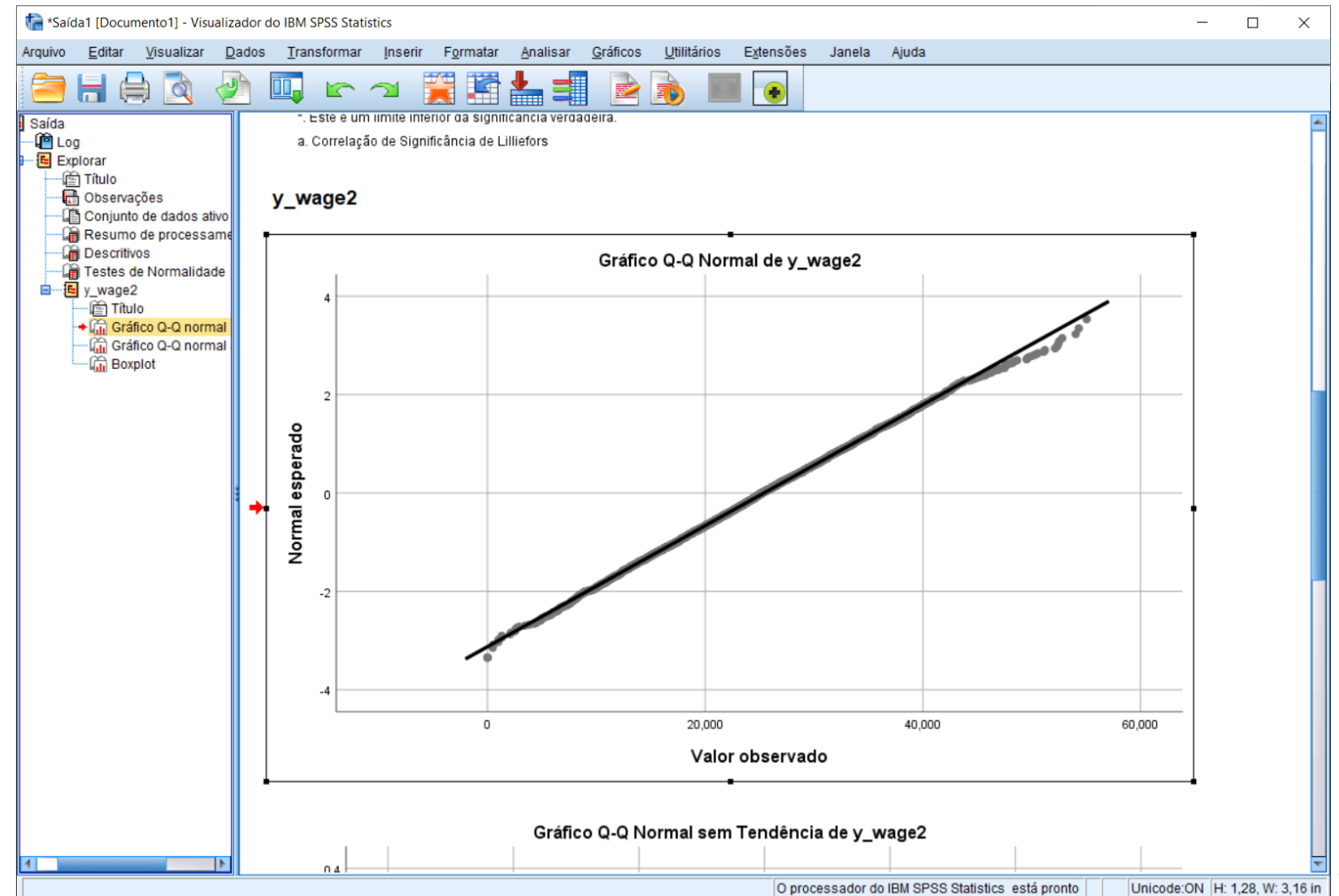
... Podemos assumir que a variável não segue uma distribuição normal.

Teste de Shapiro-Wilk

- O SPSS também oferece uma forma de visualizar se a distribuição da variável 'Idade' segue uma distribuição normal:

O gráfico Q-Q

- Neste caso, podemos assumir que a distribuição dos salários na empresa segue uma distribuição normal.



Testes de Hipóteses

A diferença entre médias (2 grupos) é significativa? (Teste de T)

Teste de T (2 Amostras Indepts.)

- Objectivo:
 - Determinar se a diferença nas médias das avaliações de homens e mulheres é estatisticamente significativa

Hipótese Nula ($H_0: \bar{X}_m = \bar{X}_f$):

“A média da avaliação dos homens é igual à média da avaliação das mulheres”

Hipótese Alternativa ($H_1: \bar{X}_m \neq \bar{X}_f$):

“A média da avaliação dos homens é diferente à média da avaliação das mulheres”

Teste de T (2 Amostras Indepts.)

- Este teste depende de um conjunto de pressupostos que convêm ser validados
 - A variável dependente é contínua;
 - A variável dependente segue uma distribuição aproximadamente normal;
 - Ausência de outliers na variável dependente
 - A variável independente é nominal, e tem apenas 2 categorias
 - As observações devem ser independentes (independência das observações)
 - Homogeneidade das variâncias (homocedasticidade)

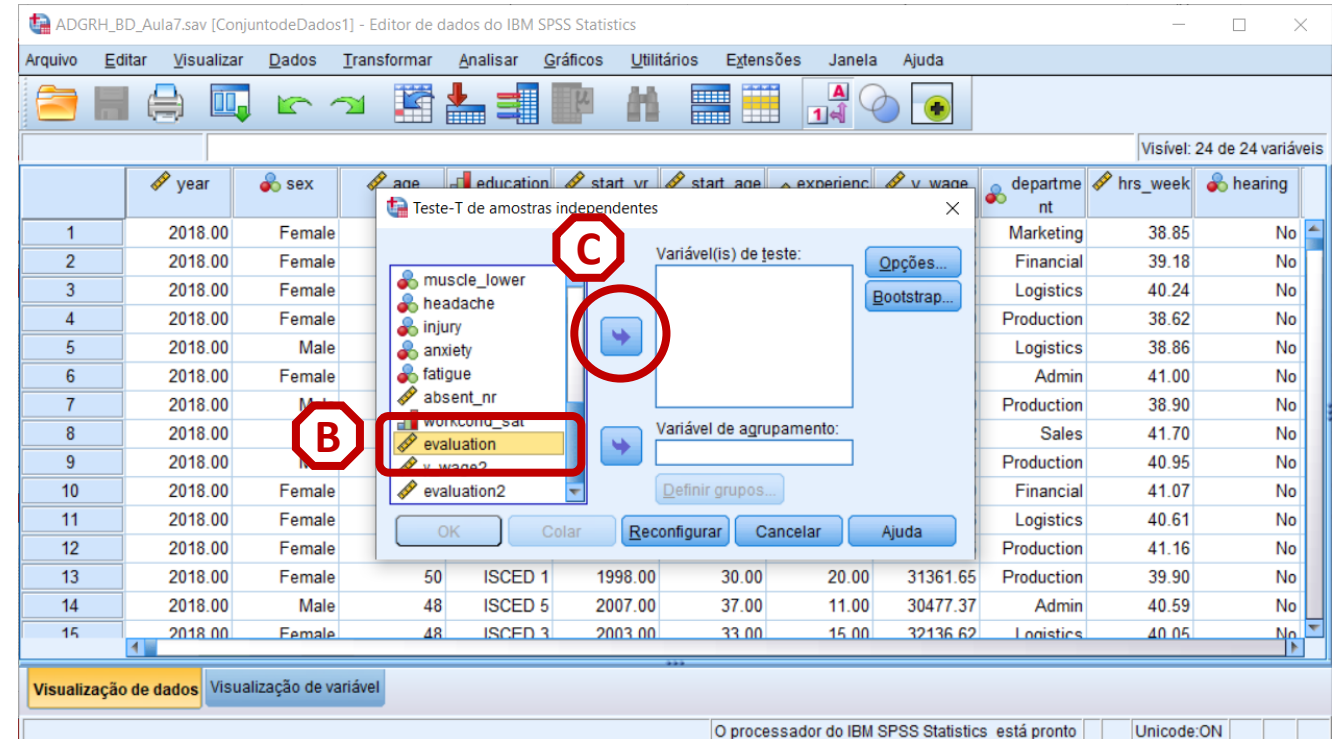
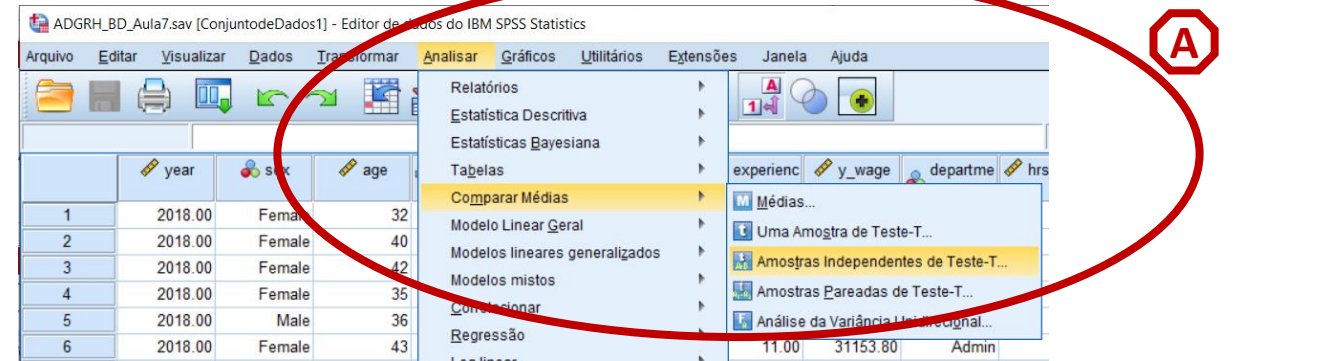
Teste de T (2 Amostras Indepts.)

- Selecionar 'Analisar / Comparar Médias' / 'Amostras independentes de Teste-T'
- Selecionar a variável 'evaluation'
- Colocar na caixa 'Variável(is) de teste'

A

B

C



Teste de T (2 Amostras Indepts.)

- Selecionar 'Analisar / Comparar Médias' / 'Amostras independentes de Teste-T'
- Selecionar a variável 'evaluation'
- Colocar na caixa 'Variável(is) de Teste'
- Selecionar a variável 'sex'
- Colocar na caixa 'Variável de agrupamento'

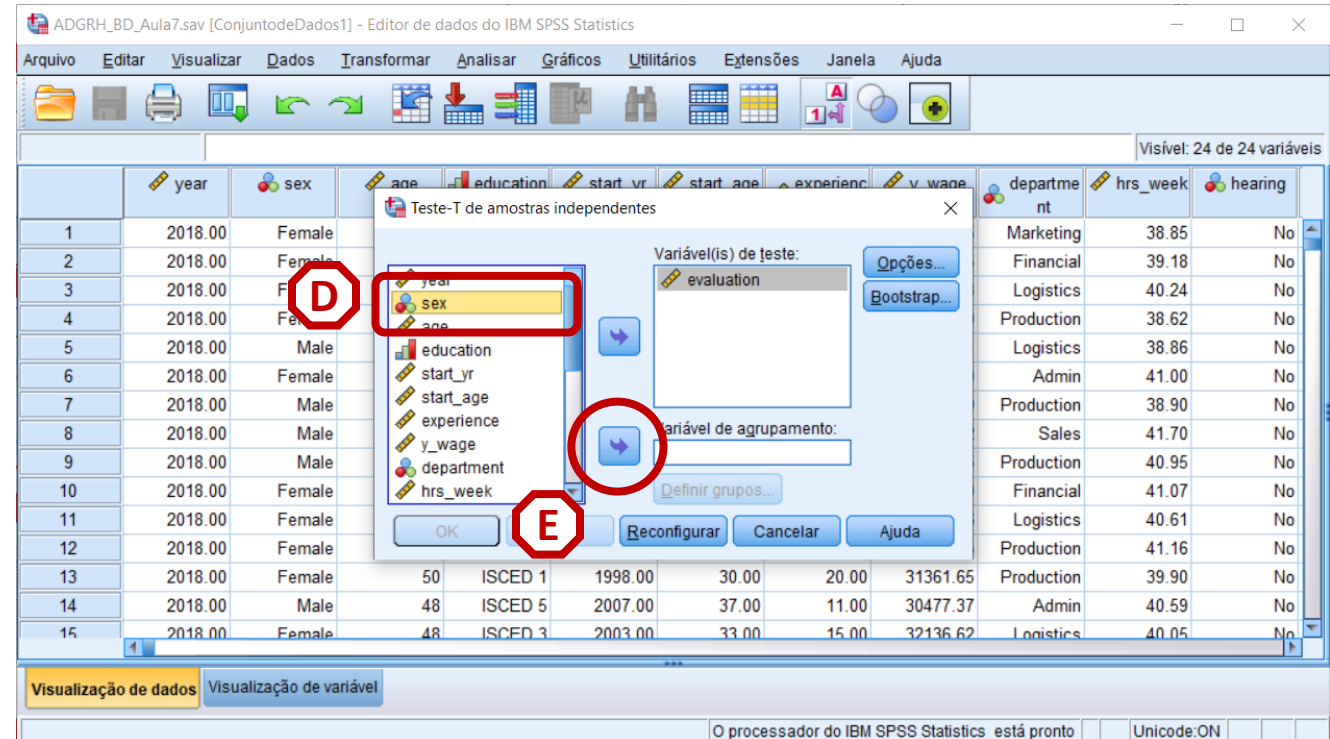
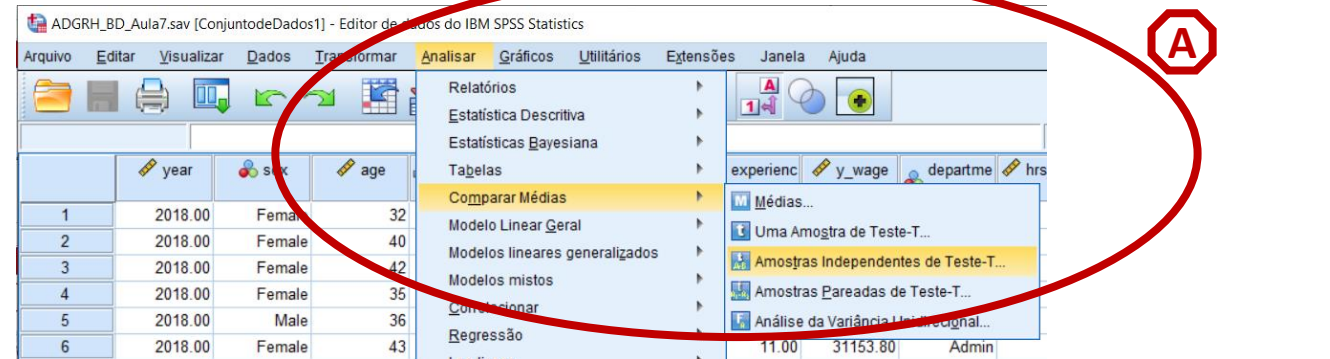
A

B

C

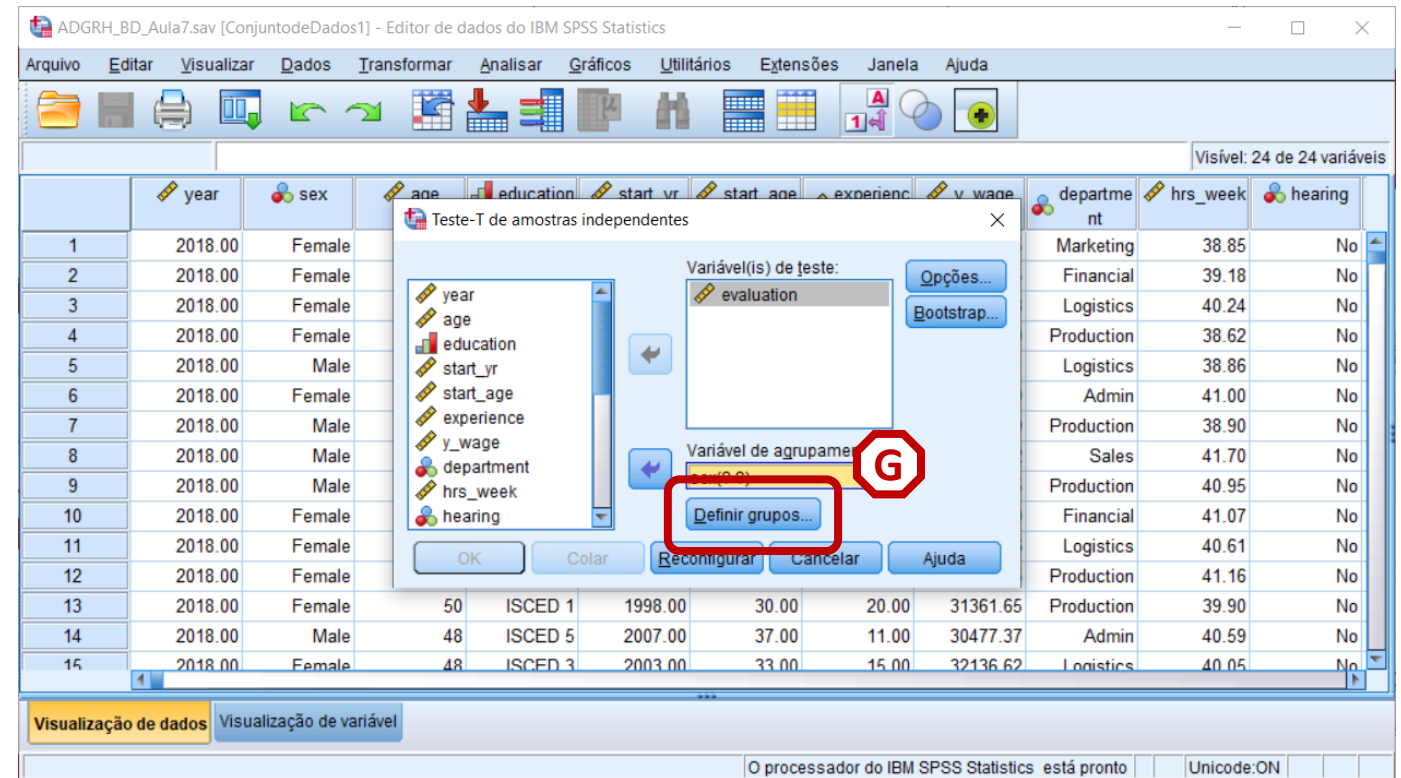
D

E



Teste de T (2 Amostras Indepts.)

- Selecionar 'Definir Grupos'



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

	year	sex	age	education	start_yr	start_age	experience	y_wage	departme nt	hrs_week	hearing
1	2018.00	Female							Marketing	38.85	No
2	2018.00	Female							Financial	39.18	No
3	2018.00	Female							Logistics	40.24	No
4	2018.00	Female							Production	38.62	No
5	2018.00	Male							Logistics	38.86	No
6	2018.00	Female							Admin	41.00	No
7	2018.00	Male							Production	38.90	No
8	2018.00	Male							Sales	41.70	No
9	2018.00	Male							Production	40.95	No
10	2018.00	Female							Financial	41.07	No
11	2018.00	Female							Logistics	40.61	No
12	2018.00	Female							Production	41.16	No
13	2018.00	Female	50	ISCED 1	1998.00	30.00	20.00	31361.65	Production	39.90	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

Teste-T de amostras independentes

Variável(is) de teste: evaluation

Variável de agrupame: department

Definir grupos...

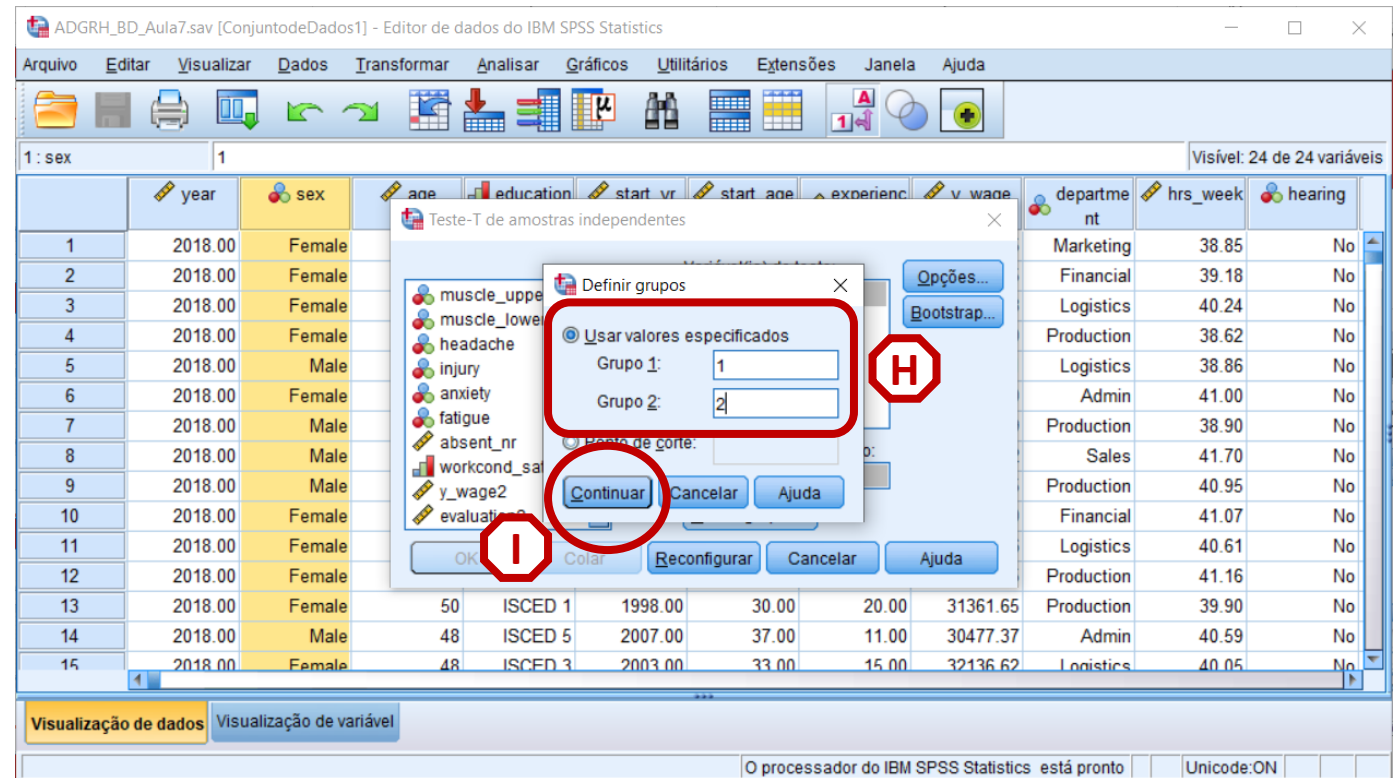
OK Color Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

Teste de T (2 Amostras Indepts.)

- Seleccionar 'Definir Grupos'
- Por os valores da variável 'sex'
- Seleccionar 'Continuar' / 'OK'

ADGRH_BD_Aula7.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: sex 1 Visível: 24 de 24 variáveis

	year	sex	age	education	start age	experience	v. wage	departme nt	hrs_week	hearing
1	2018.00	Female						Marketing	38.85	No
2	2018.00	Female						Financial	39.18	No
3	2018.00	Female						Logistics	40.24	No
4	2018.00	Female						Production	38.62	No
5	2018.00	Male						Logistics	38.86	No
6	2018.00	Female						Admin	41.00	No
7	2018.00	Male						Production	38.90	No
8	2018.00	Male						Sales	41.70	No
9	2018.00	Male						Production	40.95	No
10	2018.00	Female						Financial	41.07	No
11	2018.00	Female						Logistics	40.61	No
12	2018.00	Female						Production	41.16	No
13	2018.00	Female	50	ISCED 1	1998.00	30.00	20.00	Production	39.90	No
14	2018.00	Male	48	ISCED 5	2007.00	37.00	11.00	Admin	40.59	No
15	2018.00	Female	48	ISCED 3	2003.00	33.00	15.00	Logistics	40.05	No

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

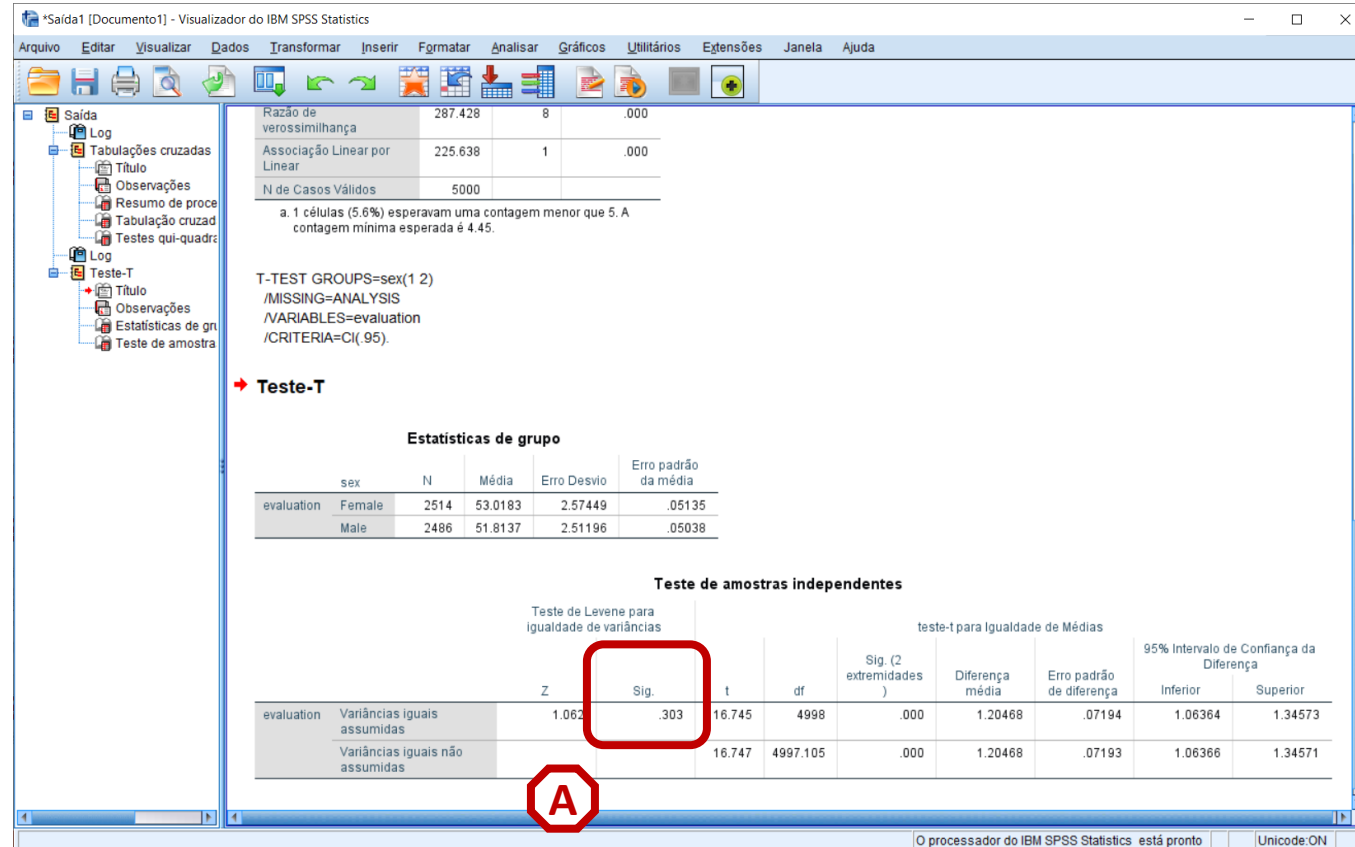
Teste de T (2 Amostras Indepts.)

- O resultado é publicado no 'Visualizador de Resultados'
- O primeiro passo é testar se o pressuposto da Homogeneidade das Variâncias se aplica.
- Para isso temos de olhar para o resultado do Teste de Levene



INTERPRETAÇÃO:

- 'Sig'. ≤ 0.05 , rejeita-se a hipótese (H_0) de que variável dependente tem a mesma variância em ambos os grupos. Aceita-se hipótese H_1
- 'Sig'. > 0.05 , não se rejeita a hipótese (H_0) de que variável dependente tem a mesma variância em ambos os grupos.



*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

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

Razão de verossimilhança 287.428 8 .000
Associação Linear por Linear 225.638 1 .000
N de Casos Válidos 5000

a. 1 células (5.6%) esperavam uma contagem menor que 5. A contagem mínima esperada é 4.45.

T-TEST GROUPS=sex(1 2)
/MISSING=ANALYSIS
/VARIABLES=evaluation
/CRITERIA=CI(.95).

→ Teste-T

Estatísticas de grupo

sex	N	Média	Erro Desvio	Erro padrão da média
evaluation Female	2514	53.0183	2.57449	.05135
Male	2486	51.8137	2.51196	.05038

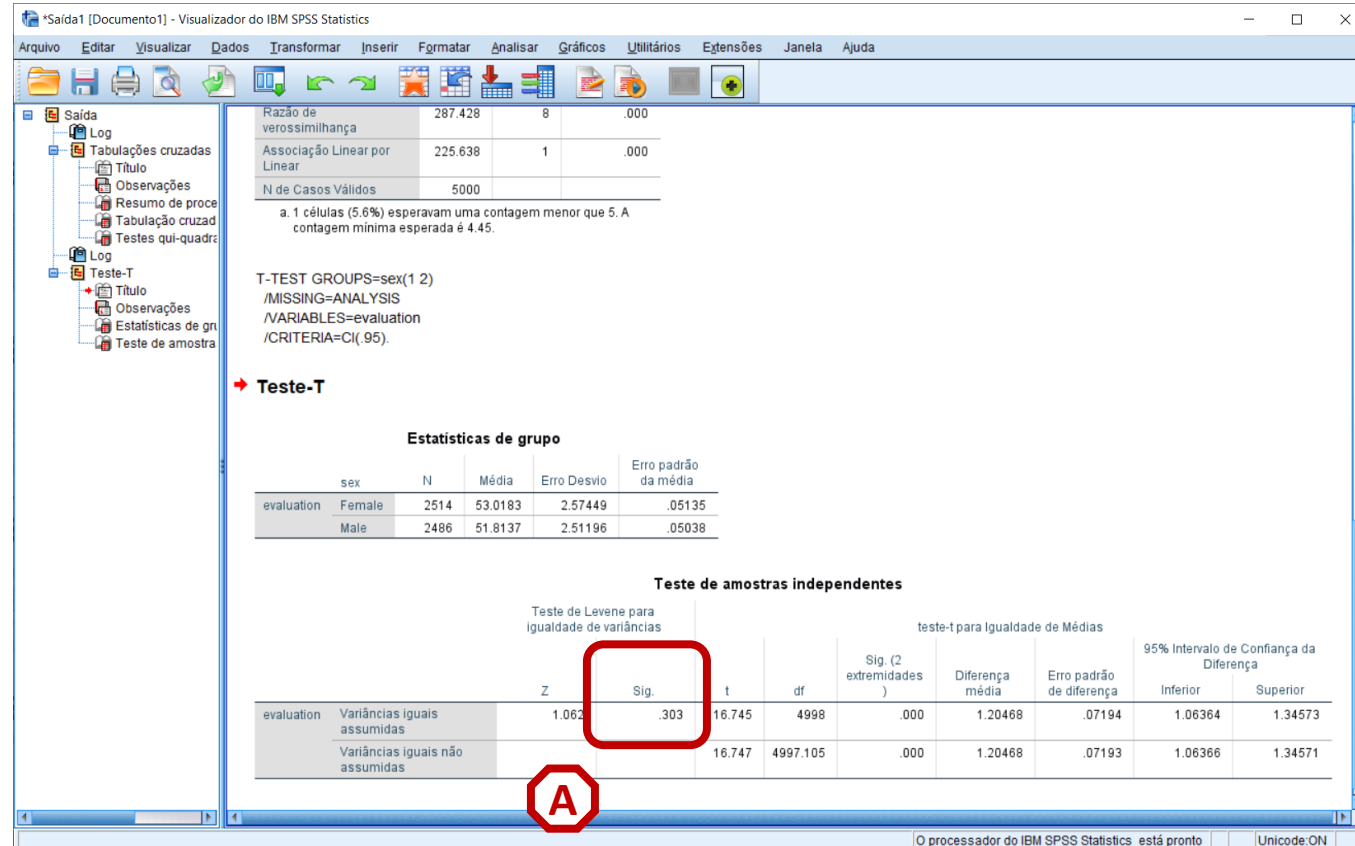
Teste de amostras independentes

		Teste de Levene para igualdade de variâncias		teste-t para Igualdade de Médias						
		Z	Sig.	t	df	Sig. (2 extremidades)	Diferença média	Erro padrão de diferença	95% Intervalo de Confiança da Diferença	
									Inferior	Superior
evaluation	Variâncias iguais assumidas	1.062	.303	16.745	4998	.000	1.20468	.07194	1.06364	1.34573
	Variâncias iguais não assumidas			16.747	4997.105	.000	1.20468	.07193	1.06366	1.34571

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

Teste de T (2 Amostras Indepts.)

- CONFIRMA-SE QUE PODEMOS ACEITAR O PRESUPOSTO DA HOMOGENEIDADE DE VARIÂNCIAS.



*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

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

Razão de verossimilhança 287.428 8 .000
Associação Linear por Linear 225.638 1 .000
N de Casos Válidos 5000

a. 1 células (5.6%) esperavam uma contagem menor que 5. A contagem mínima esperada é 4.45.

T-TEST GROUPS=sex(1 2)
/MISSING=ANALYSIS
/VARIABLES=evaluation
/CRITERIA=CI(.95).

→ Teste-T

Estatísticas de grupo

sex	N	Média	Erro Desvio	Erro padrão da média
evaluation Female	2514	53.0183	2.57449	.05135
Male	2486	51.8137	2.51196	.05038

Teste de amostras independentes

		Teste de Levene para igualdade de variâncias		teste-t para Igualdade de Médias						
		Z	Sig.	t	df	Sig. (2 extremidades)	Diferença média	Erro padrão de diferença	95% Intervalo de Confiança da Diferença	
									Inferior	Superior
evaluation	Variâncias iguais assumidas	1.062	.303	16.745	4998	.000	1.20468	.07194	1.06364	1.34573
	Variâncias iguais não assumidas			16.747	4997.105	.000	1.20468	.07193	1.06366	1.34571

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

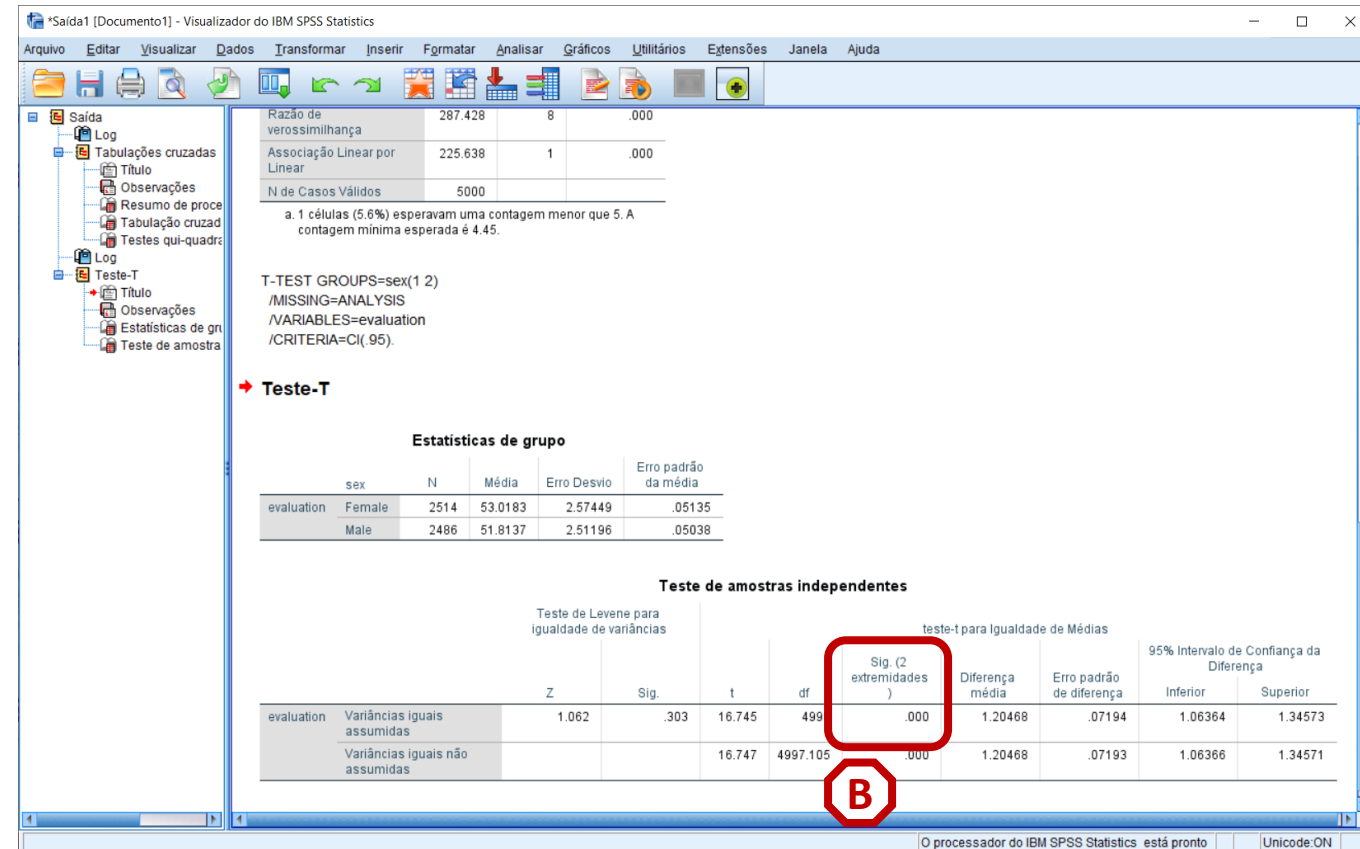
Teste de T (2 Amostras Indepts.)

- Vamos então ver o que diz o resultado do Teste de T. *

INTERPRETAÇÃO:

- 'Sig'. ≤ 0.05 , rejeita-se a hipótese (H_0) de que a média da avaliação dos homens é igual à média da avaliação das mulheres. Aceita-se hipótese H_1
- 'Sig'. > 0.05 , não se rejeita a hipótese (H_0) de que a média da avaliação dos homens é igual à média da avaliação das mulheres.

CONFIRMA-SE QUE A DIFERENÇA DAS MÉDIAS DAS AVALIAÇÕES DE HOMENS E MULHERES É ESTATÍSTICAMENTE SIGNIFICATIVA

*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

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

Razão de verossimilhança	287.428	8	.000
Associação Linear por Linear	225.638	1	.000
N de Casos Válidos	5000		

a. 1 células (5.6%) esperavam uma contagem menor que 5. A contagem mínima esperada é 4.45.

T-TEST GROUPS=sex(1 2)
/MISSING=ANALYSIS
/VARIABLES=evaluation
/CRITERIA=CI(.95).

→ Teste-T

Estatísticas de grupo

sex	N	Média	Erro Desvio	Erro padrão da média
evaluation Female	2514	53.0183	2.57449	.05135
Male	2486	51.8137	2.51196	.05038

Teste de amostras independentes

	Teste de Levene para igualdade de variâncias	teste-t para Igualdade de Médias								
		Z	Sig.	t	df	Sig. (2 extremidades)	Diferença média	Erro padrão de diferença	95% Intervalo de Confiança da Diferença	
									Inferior	Superior
evaluation	Variâncias iguais assumidas	1.062	.303	16.745	499	.000	1.20468	.07194	1.06364	1.34573
	Variâncias iguais não assumidas			16.747	4997.105	.000	1.20468	.07193	1.06366	1.34571

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



Testes de Hipóteses

Há uma relação sistemática entre as variáveis? (Qui-Quadrado)

Teste de Independência Qui-Quadrado

- Objectivo:
 - Determinar se há uma associação entre ter problemas de ansiedade e o tipo de departamento.

Hipótese Nula (H_0):

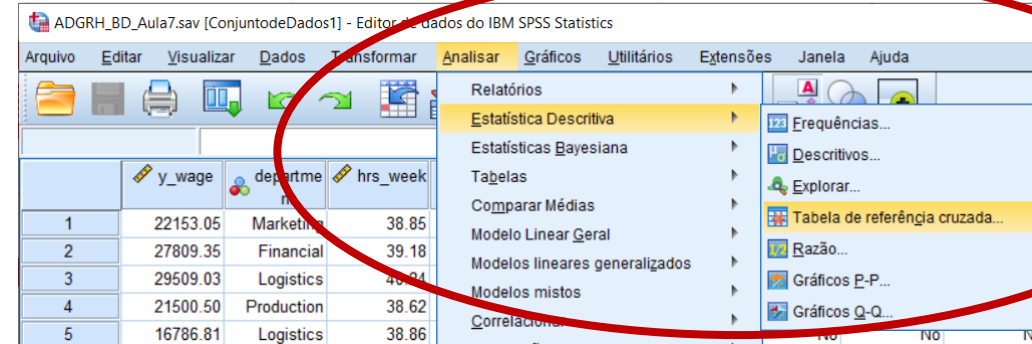
“Ter problemas de ansiedade é independente do tipo de departamento”

Hipótese Alternativa (H_1):

“Ter problemas de ansiedade não é independente do tipo de departamento”

Teste de Independência Qui-Quadrado

- Seleccionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'



Teste de Independência Qui-Quadrado

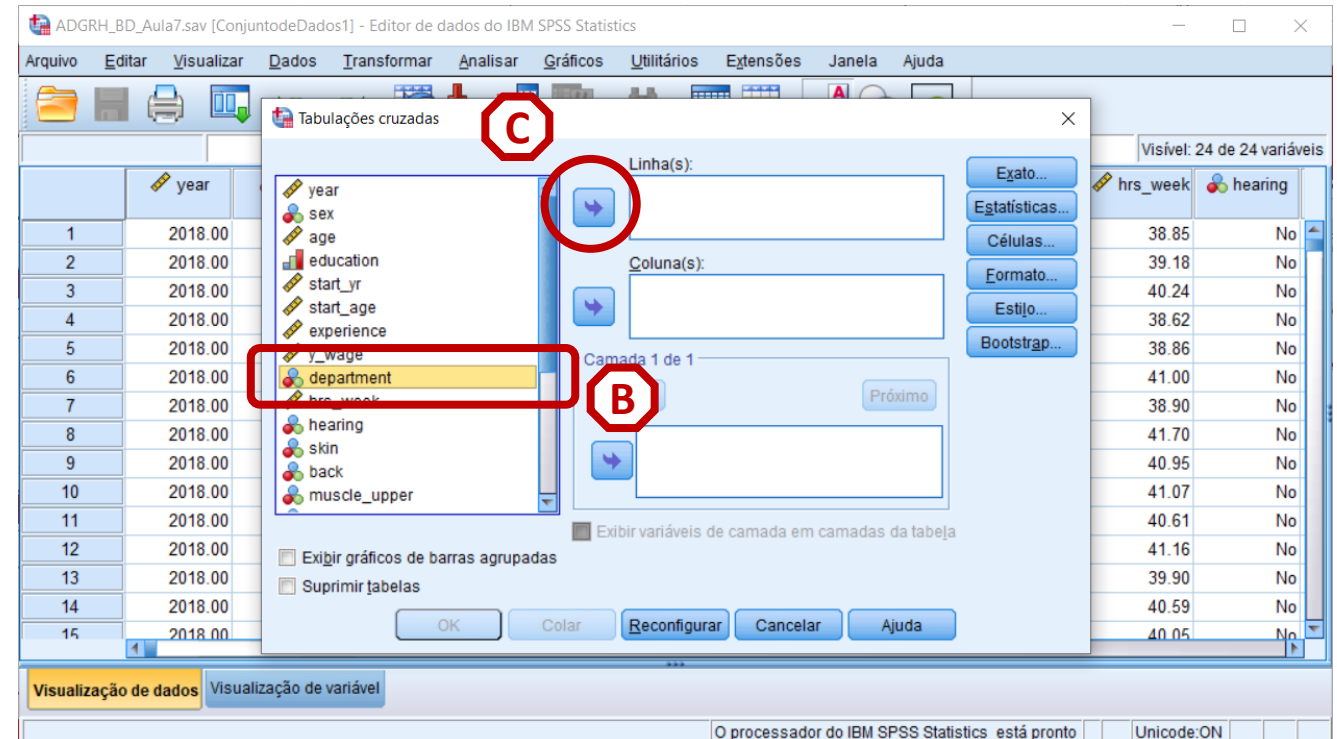
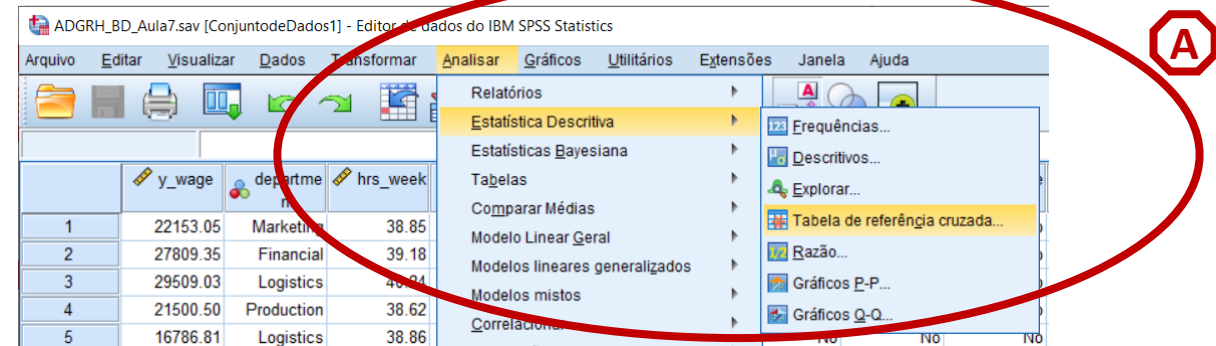
- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'
- Selecionar a variável 'department'
- Colocar na caixa 'Linha(s)'

Exercício: Colocar a variável 'anxiety' na caixa 'Coluna(s)'

A

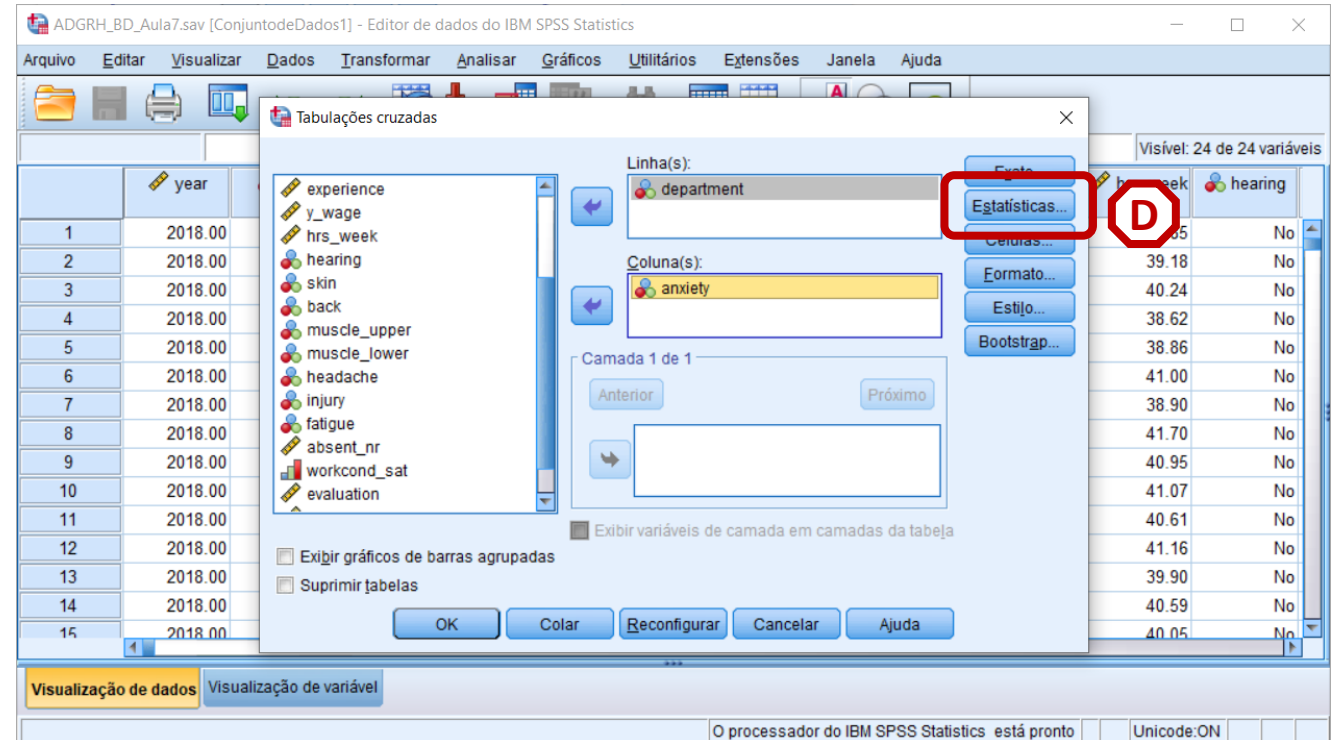
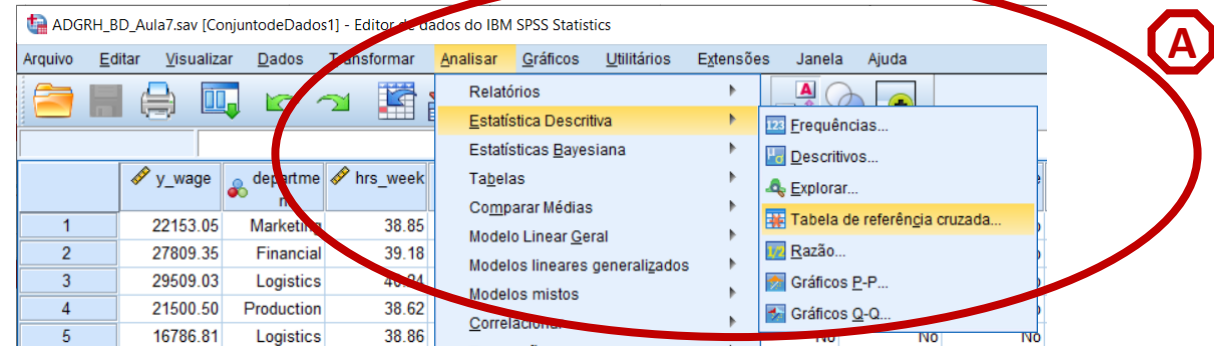
B

C



Teste de Independência Qui-Quadrado

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'
- Selecionar a variável 'department'
- Colocar na caixa 'Linha(s)'
Exercício: Colocar a variável 'anxiety' na caixa 'Coluna(s)'
- Selecionar 'Estatísticas'



Teste de Independência Qui-Quadrado

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'

- Selecionar a variável 'department'

- Colocar na caixa 'Linha(s)'

Exercício: Colocar a variável 'anxiety' na caixa 'Coluna(s)'

- Selecionar 'Estatísticas'

- Selecionar 'Qui-Quadrado'

- Selecionar 'Continuar'

A

B

C

D

E

F

ADGRH_BD_Aula7.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Tabulações cruzadas

Tabulações cruzadas: estatísticas

Qui-Quadrado

Correlações

Nominal

Coeficiente de contingência

V de Cramer e Fi

Lambda

Coeficiente de incerteza

Ordinal

Gama

d de Somers

Tau-b de Kendall

Tau-c de Kendall

Nominais por intervalo

Eta

Kappa

Rjsco

McNemar

Estatísticas de Cochran e Mantel-Haenszel

Testar a igualdade da razão da chance: 1

Continuar Cancelar Ajuda

OK Colar Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

year	hrs_week	hearing
1	2018.00	No
2	2018.00	No
3	2018.00	No
4	2018.00	No
5	2018.00	No
6	2018.00	No
7	2018.00	No
8	2018.00	No
9	2018.00	No
10	2018.00	No
11	2018.00	No
12	2018.00	No
13	2018.00	No
14	2018.00	No
15	2018.00	No

Teste de Independência Qui-Quadrado

- Selecionar 'Células'



The screenshot shows the IBM SPSS Statistics interface. The main window displays a data table with columns 'year' and 'hearing'. The 'Tabulações cruzadas' dialog box is open, showing 'department' in the 'Linha(s):' field and 'anxiety' in the 'Coluna(s):' field. The 'Células...' button is highlighted with a red box. A red octagonal icon with a white letter 'G' is overlaid on the dialog box.

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

Teste de Independência Qui-Quadrado

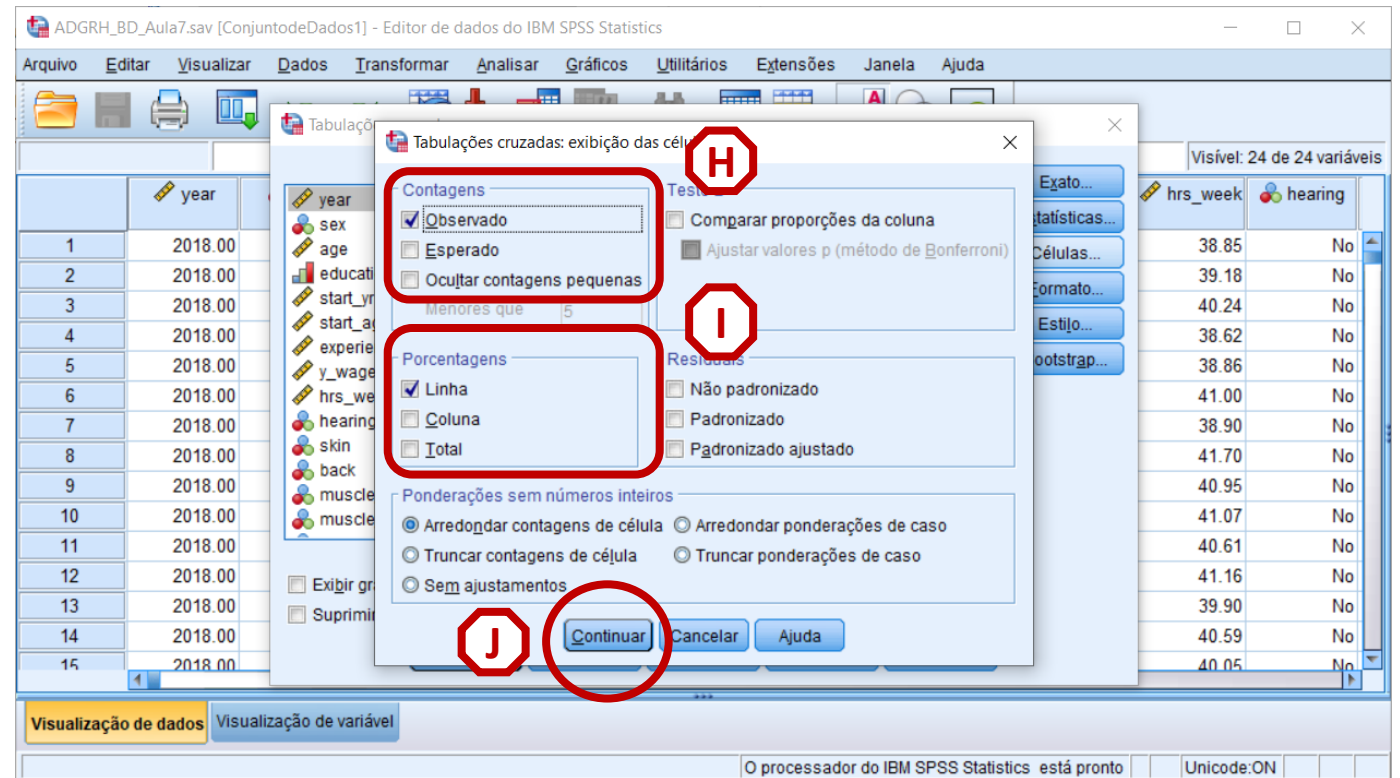
- Selecionar ‘Células’
- Selecionar ‘Contagens’ / ‘Observado’
- Selecionar ‘Porcentagens’ / ‘Linha’
- Selecionar ‘Continuar’

G

H

I

J



Teste de Independência Qui-Quadrado

- O resultado é publicado no 'Visualizador de Resultados'

INTERPRETAÇÃO:

- 'Sig'. ≤ 0.05 , rejeita-se a hipótese (H_0) de que as variáveis são independentes. Aceita-se hipótese H_1
- 'Sig'. > 0.05 , não se rejeita a hipótese (H_0) de que variáveis são independentes.

TER PROBLEMAS DE ANXIEDADE NÃO É INDEPENDENTE DO TIPO DE DEPARTAMENTO

*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

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

	% em department	99.6%	0.4%	100.0%
Sales	Contagem	450	44	494
	% em department	91.1%	8.9%	100.0%
Admin	Contagem	363	41	404
	% em department	89.9%	10.1%	100.0%
Financial	Contagem	247	35	282
	% em department	87.6%	12.4%	100.0%
Marketing	Contagem	223	27	250
	% em department	89.2%	10.8%	100.0%
IT	Contagem	199	29	228
	% em department	87.3%	12.7%	100.0%
HR	Contagem	158	22	180
	% em department	87.8%	12.2%	100.0%
Audit	Contagem	83	12	95
	% em department	87.4%	12.6%	100.0%
Total	Contagem	4766	234	5000
	% em department	95.3%	4.7%	100.0%

Testes qui-quadrado

	Valor	gl	Significância Assintótica (Bilateral)
Qui-quadrado de Pearson	279.914 ^a	8	.000
Razão de verossimilhança	287.428	8	.000
Associação Linear por Linear	225.638	1	.000
N de Casos Válidos	5000		

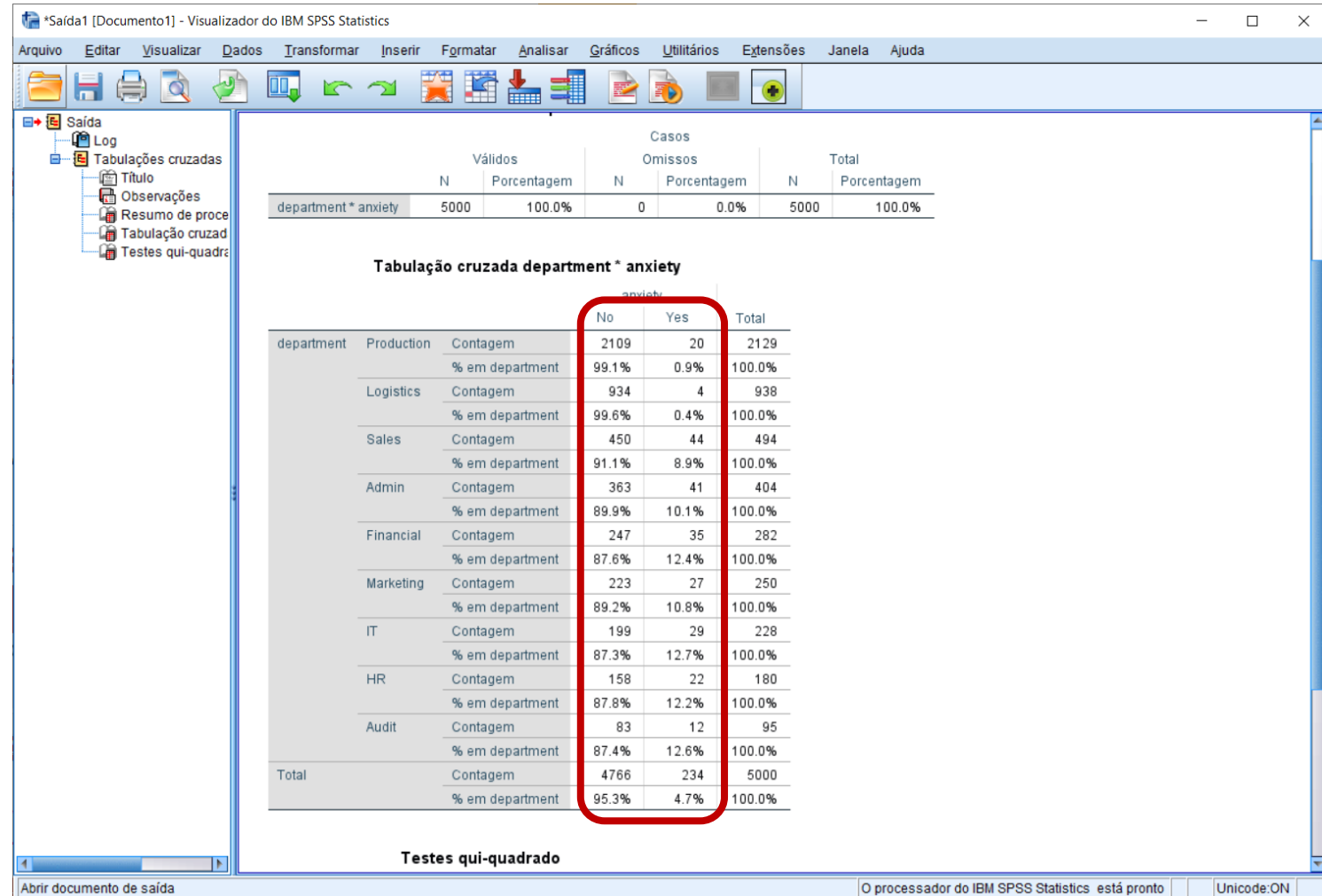
a. 1 células (5.6%) esperavam uma contagem menor que 5. A contagem mínima esperada é 4.45.

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

Teste de Independência Qui-Quadrado

- Podemos olhar para a Tabela de Frequências para melhor poder ilustrar o resultado do Teste de Independência Qui-Quadrado...

... Alguns departamentos (Sales, Financial, Marketing, IT, HR ou Audit) têm maior percentagem de trabalhadores com problemas de ansiedade que outros (production, Logistics)



Testes de Hipóteses

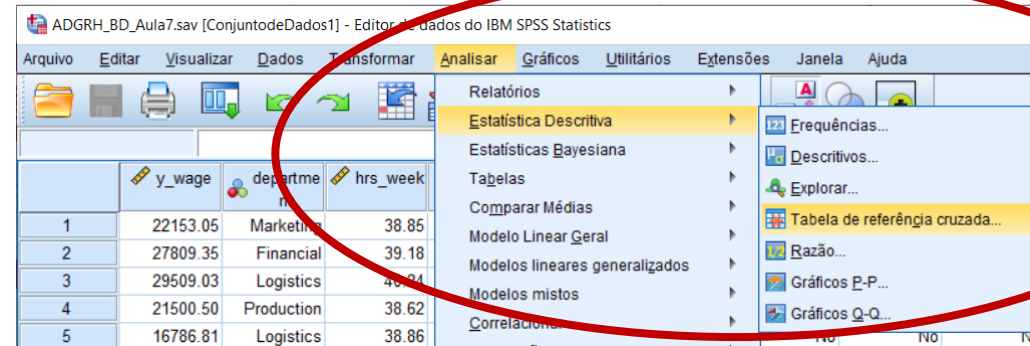
A diferença entre proporções é significativa? (Teste de Z)

Teste de Proporções (Teste de Z)

- **Objectivo:**
 - **Determinar se as diferenças na proporção de pessoas com problemas de ansiedade entre os vários departamentos são estatisticamente significativas.**

Teste de Proporções (Teste de Z)

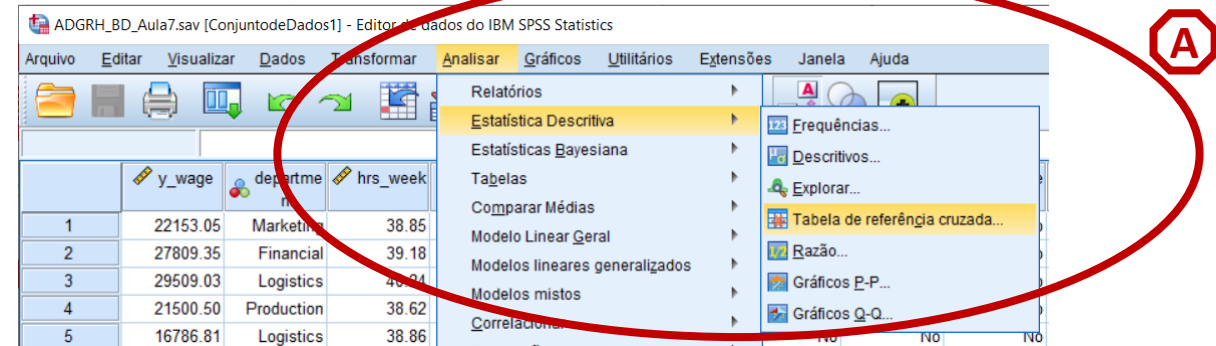
- Seleccionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'



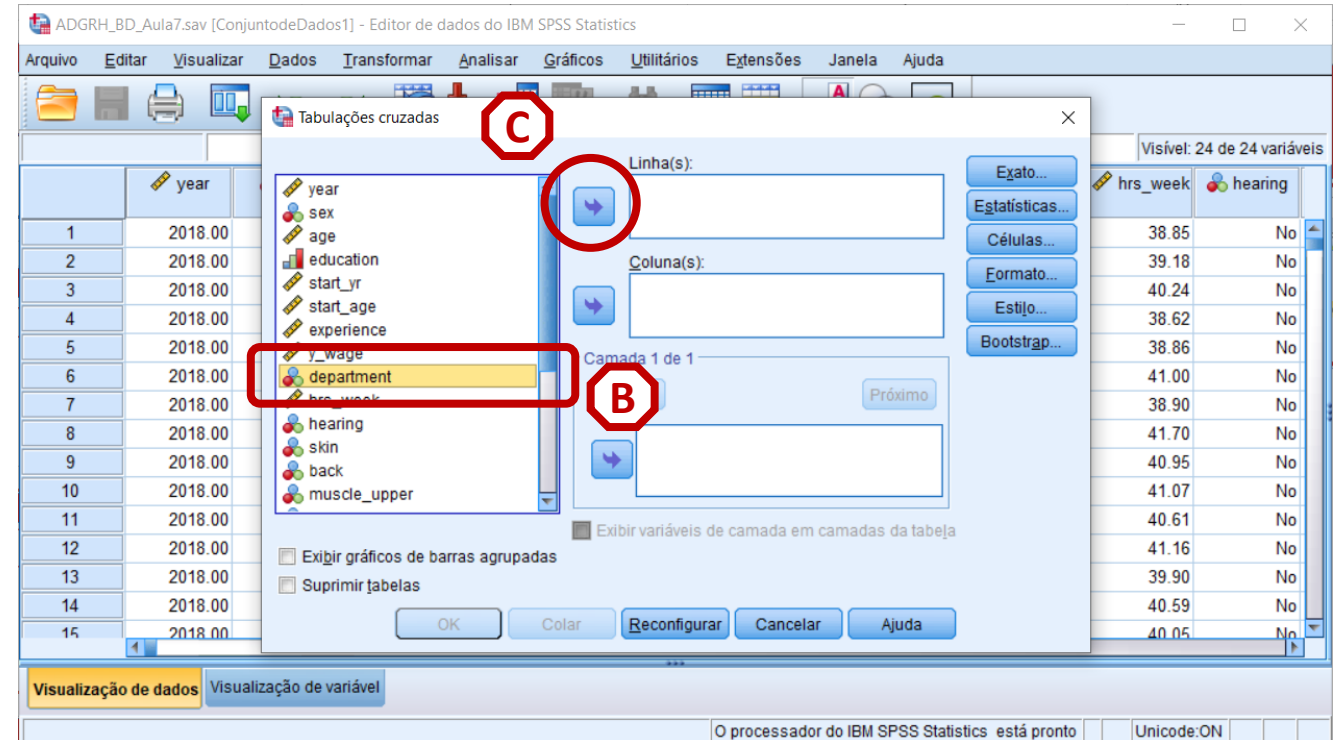
Teste de Proporções (Teste de Z)

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'
- Selecionar a variável 'department'
- Colocar na caixa 'Linha(s)'

Exercício: Colocar a variável 'anxiety' na caixa 'Coluna(s)'



A

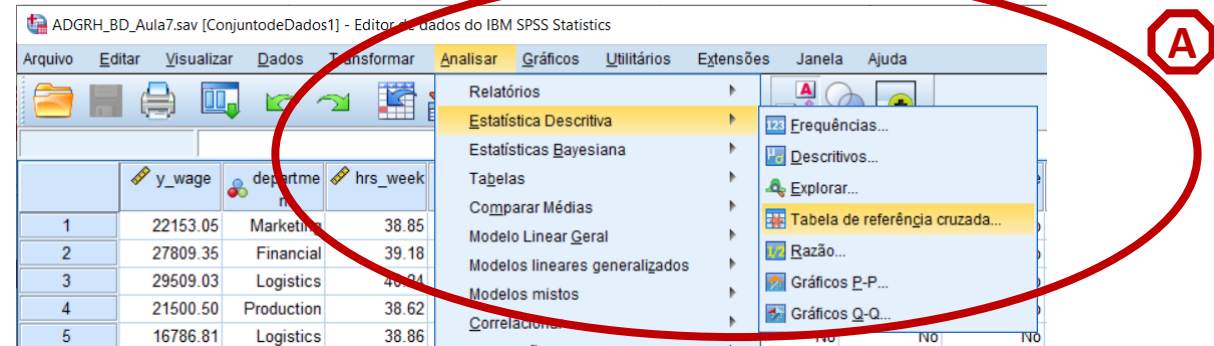


B

C

Teste de Proporções (Teste de Z)

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'



- Selecionar a variável 'department'

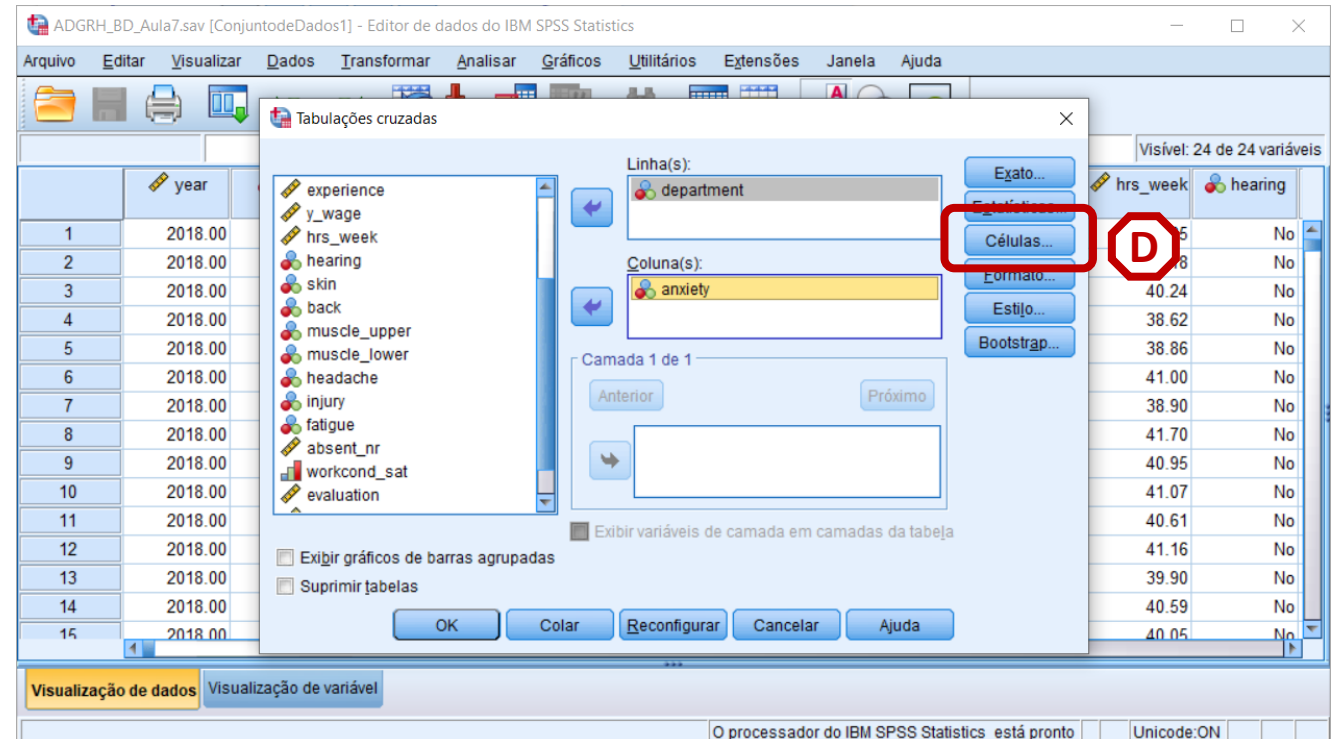


- Colocar na caixa 'Linha(s)'



Exercício: Colocar a variável 'anxiety' na caixa 'Coluna(s)'

- Selecionar 'Células'



Teste de Proporções (Teste de Z)

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'
- Selecionar a variável 'department'
- Colocar na caixa 'Linha(s)'
 - Exercício: Colocar a variável 'anxiety' na caixa 'Coluna(s)'
- Selecionar 'Células'
- Selecionar 'Comparar proporções da coluna' (e 'Ajustar valores...')
- Selecionar 'Continuar' / 'OK'

A

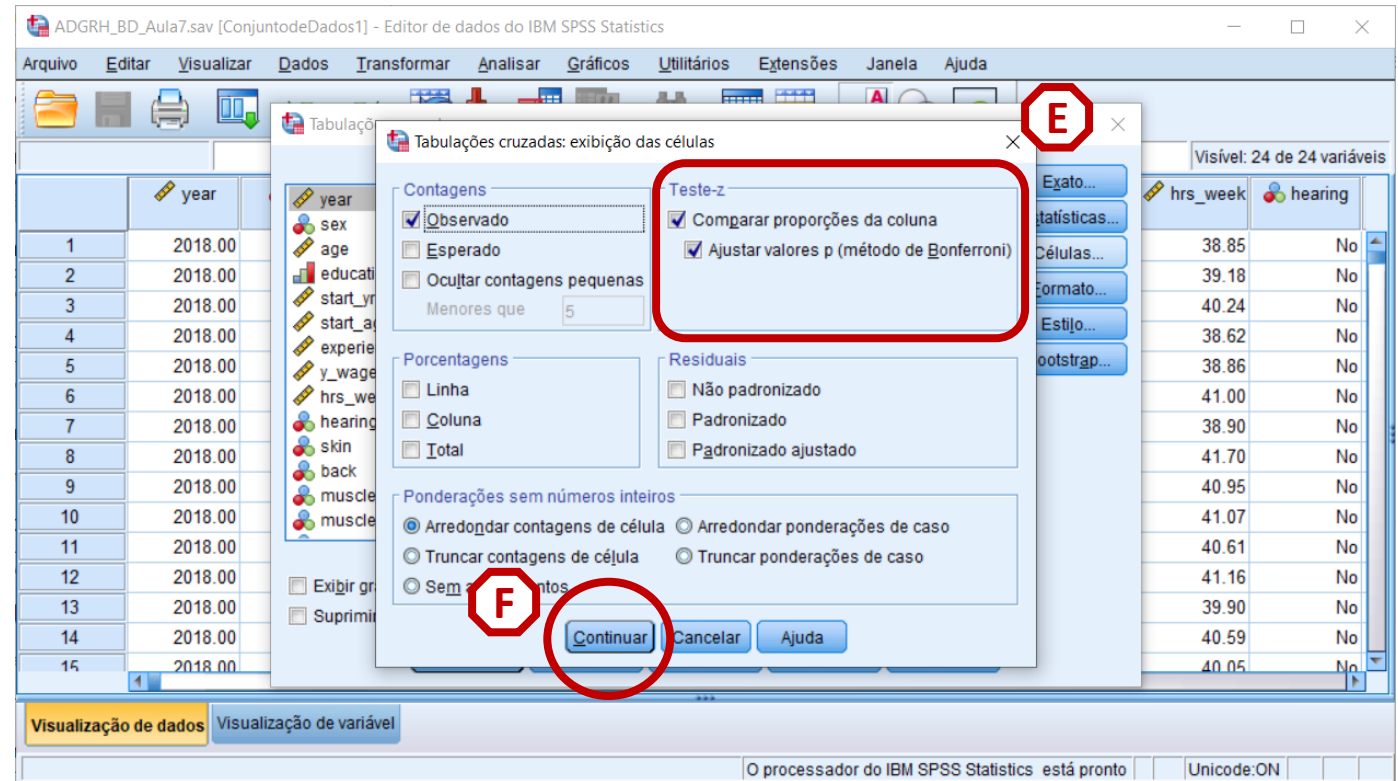
B

C

D

E

F



ADGRH_BD_Aula7.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Tabulações cruzadas: exibição das células

Contagens

- Observado
- Esperado
- Ocultar contagens pequenas

Menores que 5

Porcentagens

- Linha
- Coluna
- Total

Residuais

- Não padronizado
- Padronizado
- Padronizado ajustado

Teste-z

- Comparar proporções da coluna
- Ajustar valores p (método de Bonferroni)

Ponderações sem números inteiros


- Arredondar contagens de célula
- Truncar contagens de célula
- Sem arredondamento
- Arredondar ponderações de caso
- Truncar ponderações de caso

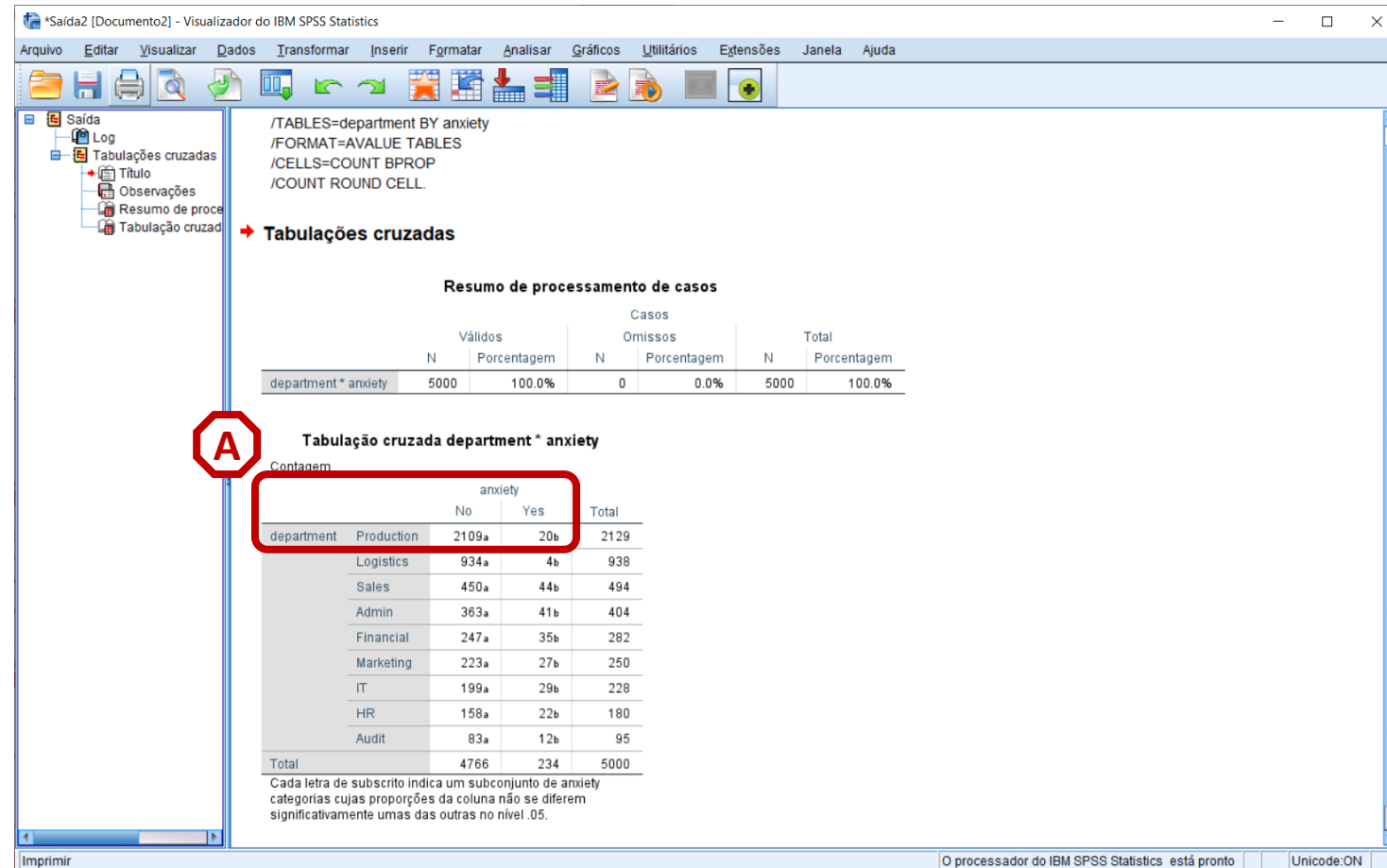
Continuar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

Teste de Proporções (Teste de Z)

- O resultado é publicado no 'Visualizador de Resultados'
- O teste atribui uma letra subscrita às categorias da variável da coluna.
- Por exemplo, para o departamento 'Production', o valor na célula 'No' tem o subscrito *a* e o valor na célula 'Yes' tem o subscrito *b*. 
- Se as colunas tiverem subscritos diferentes, isso significa que as proporções nessas células são significativamente diferentes.



Visualizador do IBM SPSS Statistics

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

Log
Tabulações cruzadas
Título
Observações
Resumo de proce
Tabulação cruzad

Tabulações cruzadas

Resumo de processamento de casos

	Válidos		Casos Omissos		Total	
	N	Porcentagem	N	Porcentagem	N	Porcentagem
department * anxiety	5000	100.0%	0	0.0%	5000	100.0%

Tabulação cruzada department * anxiety

Contagem	department	anxiety		Total
		No	Yes	
	Production	2109 ^a	20 ^b	2129
	Logistics	934 ^a	4 ^b	938
	Sales	450 ^a	44 ^b	494
	Admin	363 ^a	41 ^b	404
	Financial	247 ^a	35 ^b	282
	Marketing	223 ^a	27 ^b	250
	IT	199 ^a	29 ^b	228
	HR	158 ^a	22 ^b	180
	Audit	83 ^a	12 ^b	95
	Total	4766	234	5000

Cada letra de subscrito indica um subconjunto de anxiety categorias cujas proporções da coluna não se diferem significativamente umas das outras no nível .05.

Imprimir

O processador do IBM SPSS Statistics está pronto

Unicode:ON

Teste de Proporções (Teste de Z)

Mas a implementação deste tipo de testes não é sempre assim tão clara...

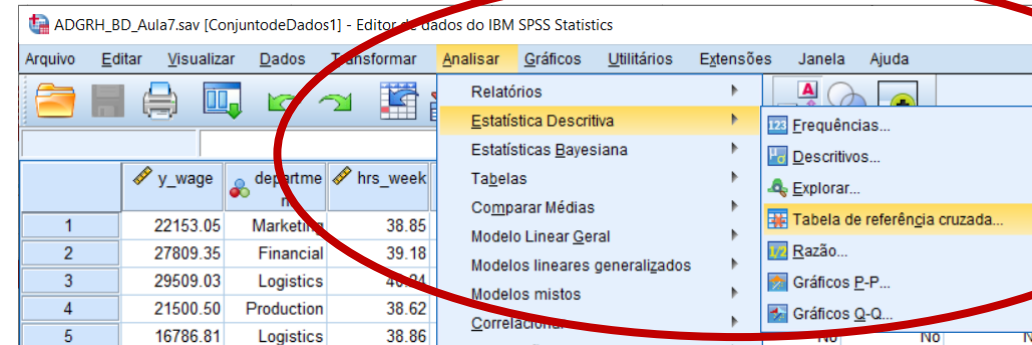
Vamos ver outro exemplo.

Teste de Proporções (Teste de Z)

- **Objectivo:**
 - **Determinar se as diferenças no grau de escolaridade dos trabalhadores entre os vários departamentos são estatisticamente significativas.**

Teste de Proporções (Teste de Z)

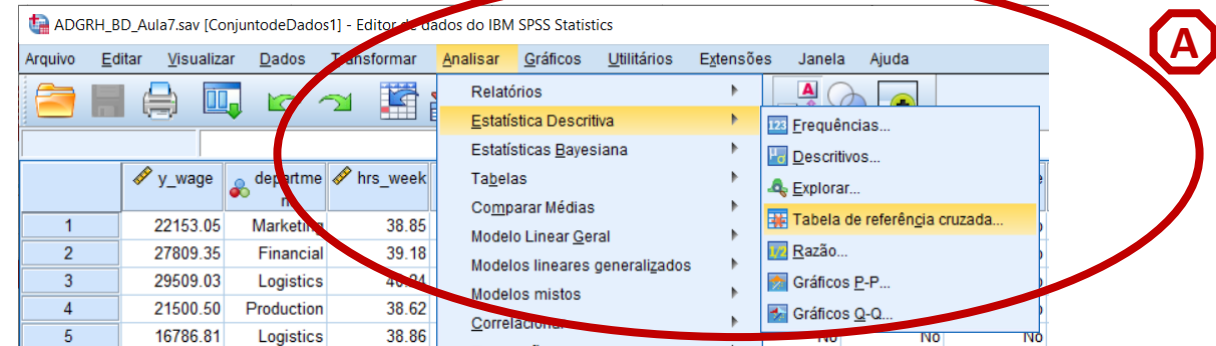
- Seleccionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'



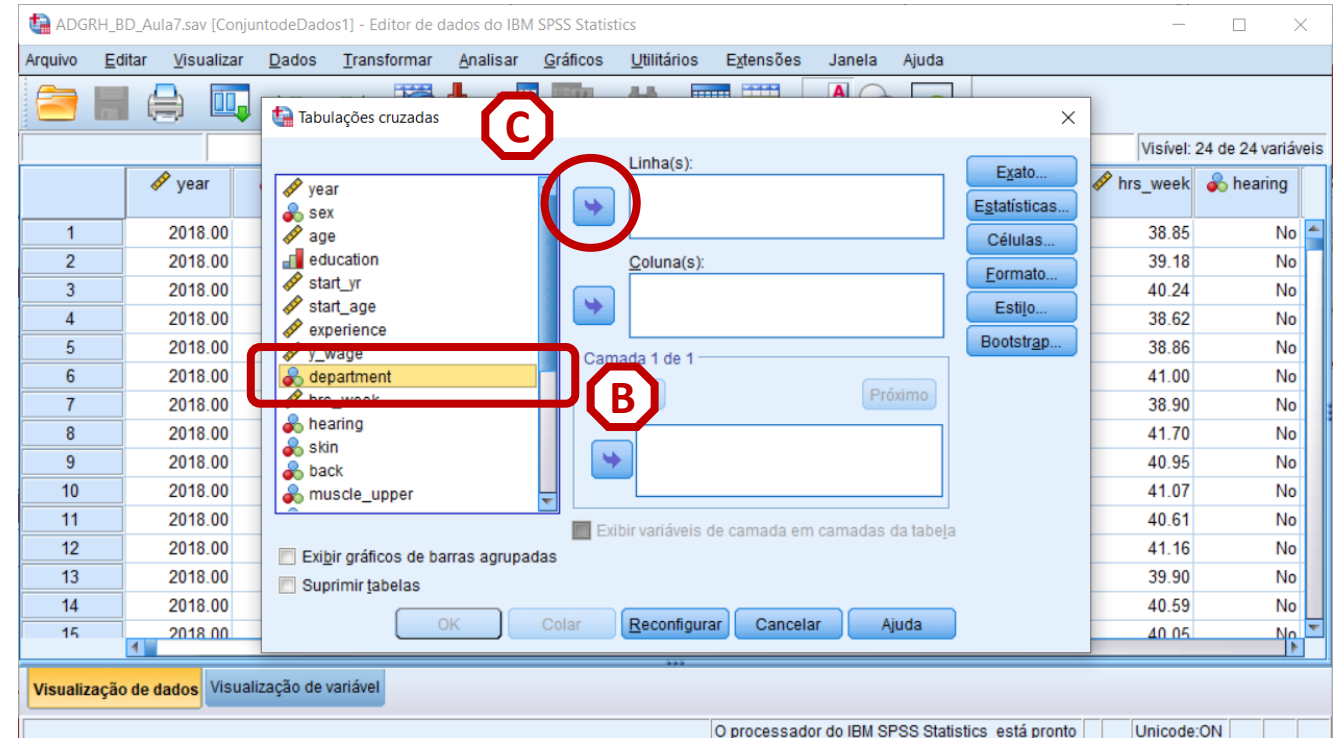
Teste de Proporções (Teste de Z)

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'
- Selecionar a variável 'department'
- Colocar na caixa 'Linha(s)'

Exercício: Colocar a variável 'education' na caixa 'Coluna(s)'



A

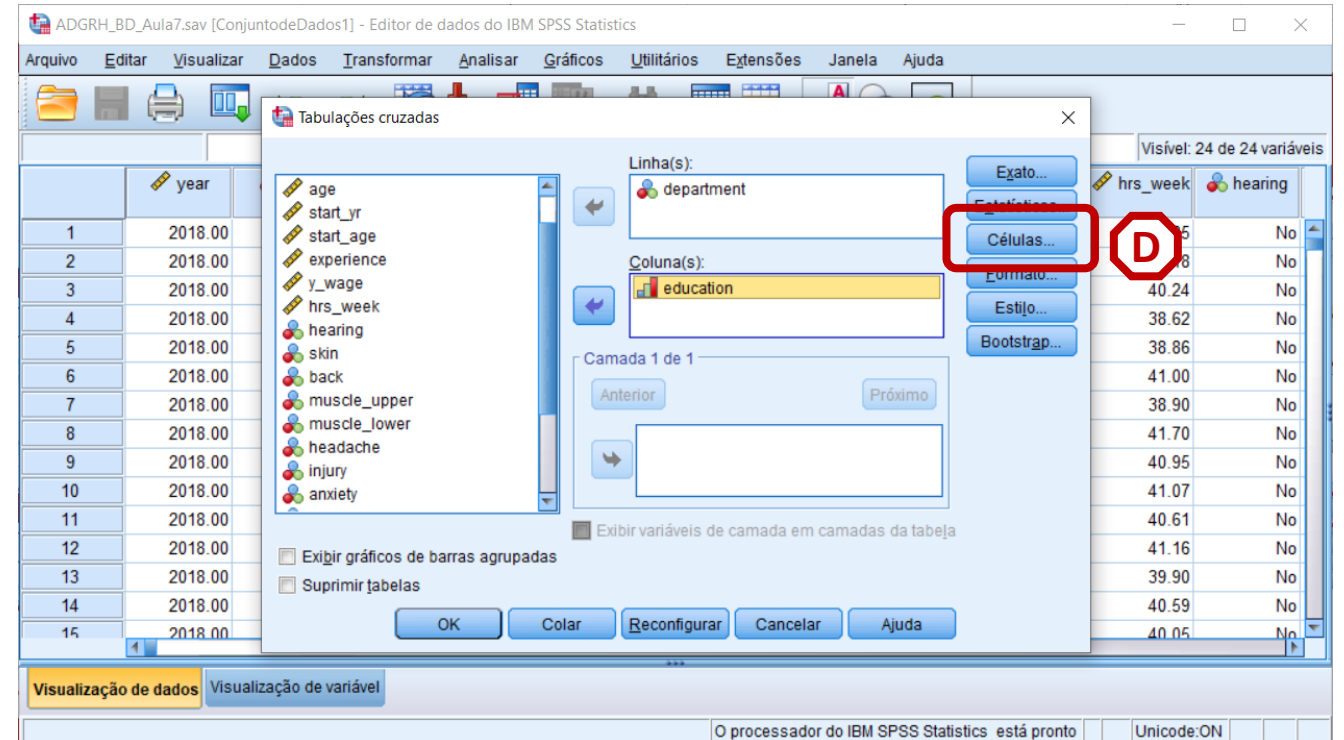
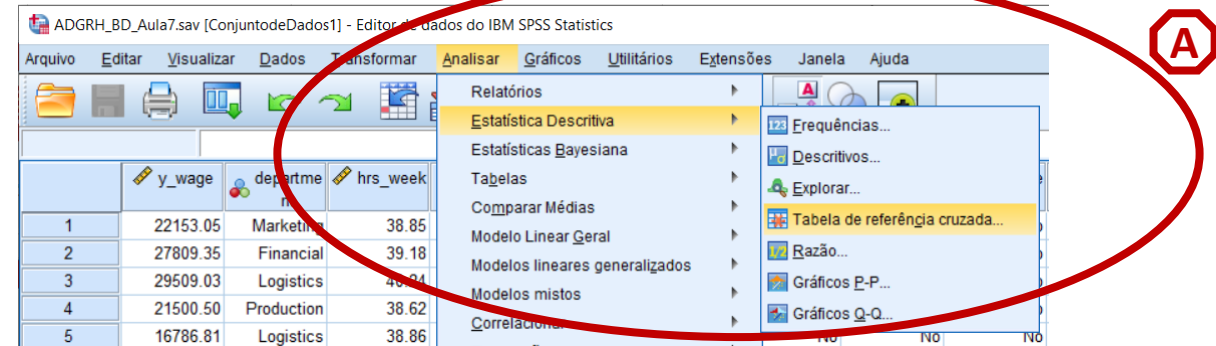


B

C

Teste de Proporções (Teste de Z)

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'
- Selecionar a variável 'department'
- Colocar na caixa 'Linha(s)'
Exercício: Colocar a variável 'education' na caixa 'Coluna(s)'
- Selecionar 'Células'



Teste de Proporções (Teste de Z)

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Tabela de referência cruzada'
- Selecionar a variável 'department'
- Colocar na caixa 'Linha(s)'
Exercício: Colocar a variável 'education' na caixa 'Coluna(s)'
- Selecionar 'Células'
- Selecionar 'Comparar proporções da coluna' (e 'Ajustar valores...')
- Selecionar 'Continuar' / 'OK'

A

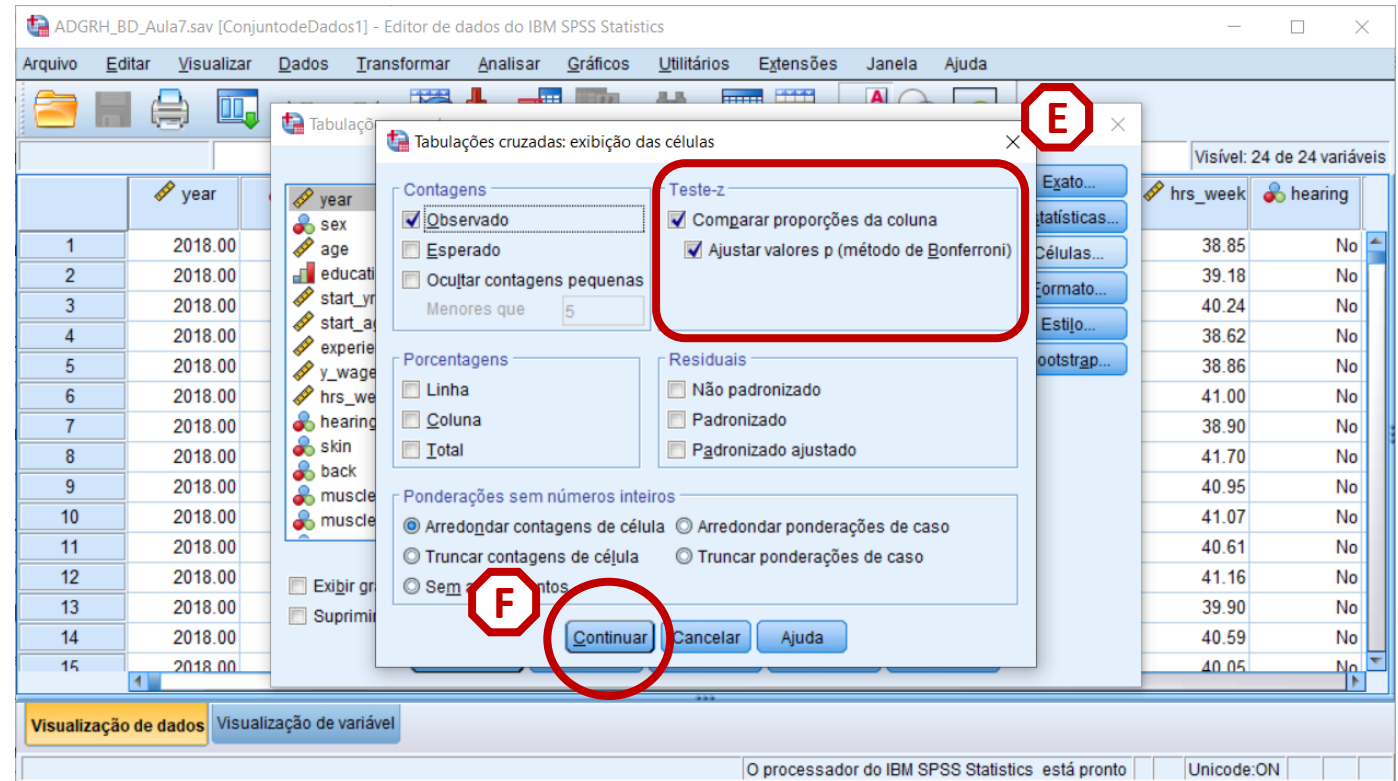
B

C

D

E

F



ADGRH_BD_Aula7.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

Tabulações cruzadas: exibição das células

Contagens

- Observado
- Esperado
- Ocultar contagens pequenas

Menores que 5

Porcentagens

- Linha
- Coluna
- Total

Residuais

- Não padronizado
- Padronizado
- Padronizado ajustado

Teste-z

- Comparar proporções da coluna
- Ajustar valores p (método de Bonferroni)

Ponderações sem números inteiros

- Arredondar contagens de célula
- Truncar contagens de célula
- Sem arredondamento
- Arredondar ponderações de caso
- Truncar ponderações de caso

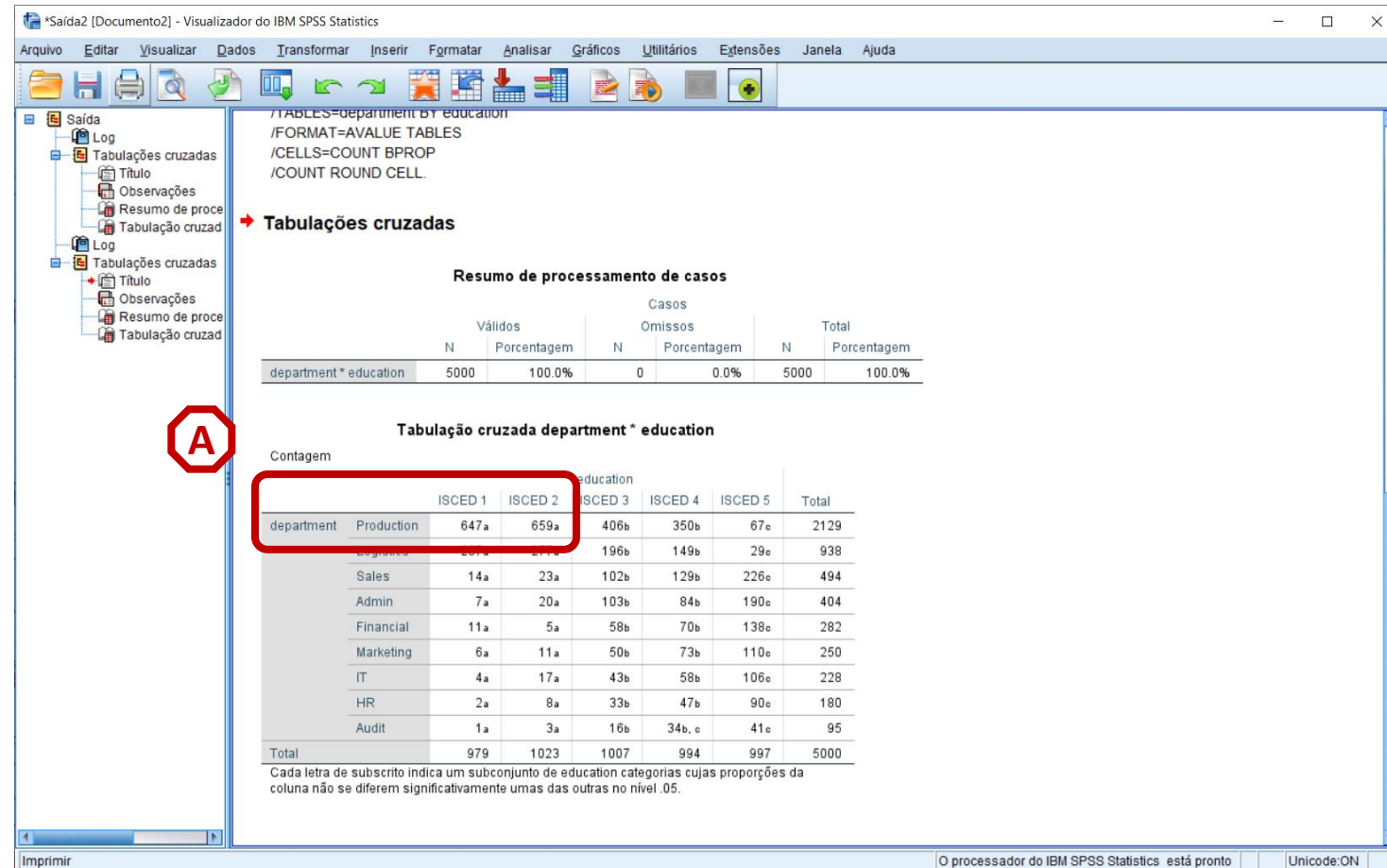
Continuar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

Teste de Proporções (Teste de Z)

- O resultado é publicado no 'Visualizador de Resultados'
- Olhando para o departamento de 'Production'...
- ... os valores nas categorias 'ISCED1' e 'ISCED2' têm o mesmo índice α , o que significa que não podem ser distinguidos uns dos outros. A
- Por outro lado, os subscritos nas categorias 'ISCED4' e 'ISCED5' diferem uns dos outros.
- O que sugere que apenas as percentagens na categoria 'ISCED5' são significativamente diferentes (excepto no departamento de 'Audit').



Visualizador do IBM SPSS Statistics

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

Log
Tabulações cruzadas
Título
Observações
Resumo de proce
Tabulação cruzad
Log
Tabulações cruzadas
Título
Observações
Resumo de proce
Tabulação cruzad

```

/!TABLES=department BY education
/FORMAT=AVALUE TABLES
/CELLS=COUNT BPROP
/COUNT ROUND CELL.
  
```

→ **Tabulações cruzadas**

Resumo de processamento de casos

	Válidos		Casos Omissos		Total	
	N	Porcentagem	N	Porcentagem	N	Porcentagem
department * education	5000	100.0%	0	0.0%	5000	100.0%

Tabulação cruzada department * education

Contagem		education					Total
		ISCED 1	ISCED 2	ISCED 3	ISCED 4	ISCED 5	
department	Production	647 ^a	659 ^a	406 ^b	350 ^b	67 ^e	2129
	Engineering	159 ^a	177 ^a	196 ^b	149 ^b	29 ^e	938
	Sales	14 ^a	23 ^a	102 ^b	129 ^b	226 ^e	494
	Admin	7 ^a	20 ^a	103 ^b	84 ^b	190 ^e	404
	Financial	11 ^a	5 ^a	58 ^b	70 ^b	138 ^e	282
	Marketing	6 ^a	11 ^a	50 ^b	73 ^b	110 ^e	250
	IT	4 ^a	17 ^a	43 ^b	58 ^b	106 ^e	228
	HR	2 ^a	8 ^a	33 ^b	47 ^b	90 ^e	180
	Audit	1 ^a	3 ^a	16 ^b	34 ^{b, c}	41 ^e	95
Total		979	1023	1007	994	997	5000

Nota: Cada letra de subscrito indica um subconjunto de education categorias cujas proporções da coluna não se diferem significativamente umas das outras no nível .05.

Imprimir

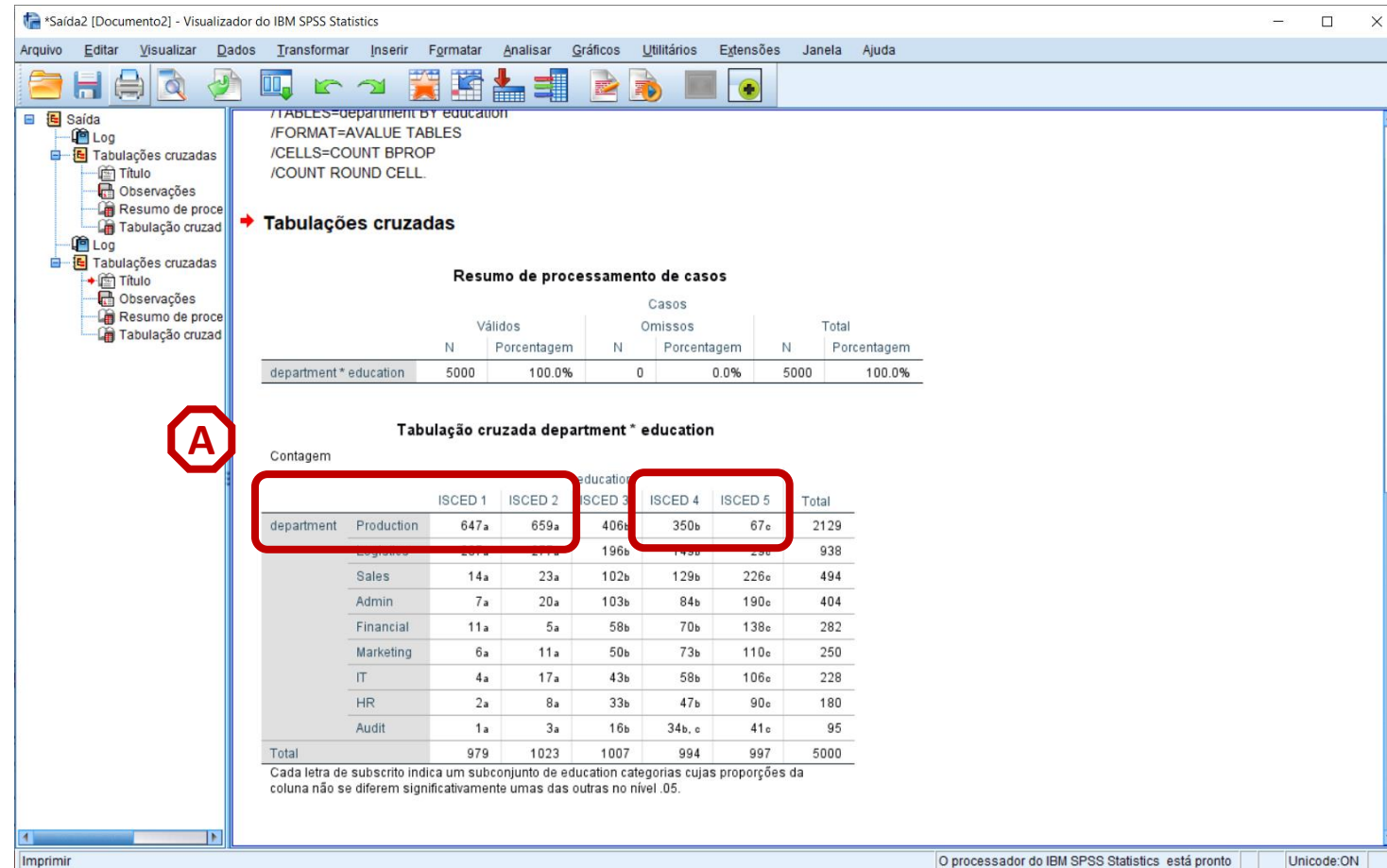
O processador do IBM SPSS Statistics está pronto Unicode:ON

Teste de Proporções (Teste de Z)

- O resultado é publicado no 'Visualizador de Resultados'
- Olhando para o departamento de 'Production'...
- ... os valores nas categorias 'ISCED1' e 'ISCED2' têm o mesmo índice α , o que significa que não podem ser distinguidos uns dos outros.
- Por outro lado, os subscritos nas categorias 'ISCED4' e 'ISCED5' diferem uns dos outros.
- O que sugere que apenas as percentagens na categoria 'ISCED5' são significativamente diferentes de todas as outras (excepto no departamento de 'Audit').

A

A



Visualizador do IBM SPSS Statistics

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

Log
Tabulações cruzadas
Título
Observações
Resumo de proce
Tabulação cruzad
Log
Tabulações cruzadas
Título
Observações
Resumo de proce
Tabulação cruzad

→ Tabulações cruzadas

Resumo de processamento de casos

	Válidos		Casos Omissos		Total	
	N	Porcentagem	N	Porcentagem	N	Porcentagem
department * education	5000	100.0%	0	0.0%	5000	100.0%

Tabulação cruzada department * education

Contagem		education					Total
		ISCED 1	ISCED 2	ISCED 3	ISCED 4	ISCED 5	
department	Production	647 ^a	659 ^a	406 ^b	350 ^b	67 ^e	2129
	Engineering	255 ^b	277 ^b	196 ^b	143 ^b	230 ^c	938
	Sales	14 ^a	23 ^a	102 ^b	129 ^b	226 ^c	494
	Admin	7 ^a	20 ^a	103 ^b	84 ^b	190 ^c	404
	Financial	11 ^a	5 ^a	58 ^b	70 ^b	138 ^c	282
	Marketing	6 ^a	11 ^a	50 ^b	73 ^b	110 ^c	250
	IT	4 ^a	17 ^a	43 ^b	58 ^b	106 ^c	228
	HR	2 ^a	8 ^a	33 ^b	47 ^b	90 ^c	180
	Audit	1 ^a	3 ^a	16 ^b	34 ^{b, c}	41 ^c	95
Total		979	1023	1007	994	997	5000

Cada letra de subscrito indica um subconjunto de education categorias cujas proporções da coluna não se diferem significativamente umas das outras no nível .05.

Imprimir

O processador do IBM SPSS Statistics está pronto Unicode:ON

- **Resultados possíveis de um teste de hipótese**

	A HIPÓTESE NULA É VERDADEIRA	A HIPÓTESE NULA É FALSA
REJEITA-SE A HIPÓTESE NULA	Erro de Tipo I	Decisão Correta
NÃO SE REJEITA A HIPÓTESE NULA	Decisão Correta	Erro de Tipo II

- Resultados possíveis de um teste de hipótese

	A HIPÓTESE NULA É VERDADEIRA	A HIPÓTESE NULA É FALSA
REJEITA-SE A HIPÓTESE NULA	Erro de Tipo I (α)	Decisão Correta
NÃO SE REJEITA A HIPÓTESE NULA	Decisão Correta	Erro de Tipo II

- Erro de Tipo 1
 - Representado pelo símbolo α (alpha)
 - É igual ao 'Nível de Significância' que atribuímos ao teste quando escolhemos o Grau de Confiança.

Level of Significance	Confidence Level (%)
$\alpha=0.001$	99.9
$\alpha=0.01$	99
$\alpha=0.05$	95
$\alpha=0.1$	90

- Resultados possíveis de um teste de hipótese

	A HIPÓTESE NULA É VERDADEIRA	A HIPÓTESE NULA É FALSA
REJEITA-SE A HIPÓTESE NULA	Erro de Tipo I (α)	Decisão Correta
NÃO SE REJEITA A HIPÓTESE NULA	Decisão Correta	Erro de Tipo II (β)

- Erro de Tipo II
 - Representado pelo símbolo β (Beta)
 - Ao determinarmos o valor de β , podemos calcular a 'Potência Estatística' do teste, que mede a probabilidade de rejeitar a hipótese nula quando a hipótese nula é falsa (= $1 - \beta$).

ANÁLISE DE DADOS EM GRH

Aula 8: Formulação e Teste de Hipóteses (II)

Por hoje é tudo.

Até à próxima aula!