



# Engenharia de Dados Experimentais

## SPSS – Análise de Dados – Estatística Descritiva

*Raquel Guiné*



# Conteúdo

1. Frequências
2. Desritivos
3. Explorar
4. Tabelas de Contingência (Cross Tabs) e Teste do  $\chi^2$



# 1. FREQUÊNCIAS

- Comando **Frequencies** – para fazer análises univariadas (1 só variável), quer sejam qualitativas ou quantitativas (embora nós usemos mais para variáveis qualitativas)
- Ex: queremos saber a % de pessoas do sexo Feminino e do sexo Masculino, e também a distribuição em % nos diferentes anos de escolaridade.
- Fazer: **Analyze / Descriptive Statistics / Frequencies**
- Variáveis: **V1.4 (Sexo)** e **V1.5 (Ano de escolaridade)**
- Fazer **OK**.
- Ver os **OUTPUTS**: Aparecem duas tabelas de % absolutas e acumuladas. Desses dados o que interessa é apenas N e %.
- Construo novas tabelas para reportar os dados e tirar conclusões

## Sexo

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Feminino	426	50,0	50,1	50,1
	Masculino	425	49,9	49,9	100,0
	Total	851	99,9	100,0	
Missing	System	1	,1		
Total		852	100,0		

**Tabela 1.** Distribuição de inquiridos em função do sexo

Sexo	N	%
Feminino	426	50,1
Masculino	425	49,9
Totais	851	100,00

- Dos 852 participantes, um não indicou qual o seu sexo. Dos restantes, verificou-se que 50,1% eram do sexo Feminino e 49,9% do sexo Masculino.

**Tabela 2.** Distribuição de inquiridos em função do ano de escolaridade

Ano de Escolaridade	N	%
5º	187	22,0
6º	185	21,8
7º	185	21,8
8º	136	16,0
9º	157	18,5
Totais	850	100,00

- Dos 852 participantes, dois não indicaram o ano que frequentam. Dos restantes, verificou-se a maioria, 22,0% se encontravam no 5º ano.

- Também se pode usar o FREQUENCIES para variáveis quantitativas:
- Ex: **Analyze / Descriptive Statistics / Frequencies**
- Variáveis: **V1.1. (Idade)**
- **OK.**

*Só que neste caso dá uma tabela muito complexa, que não é fácil trabalhar ou entender e não dá muita informação (p.ex., não dá médias).*

*Neste caso, como a amostra tinha uma gama de idades muito limitada (entre 10 e 18 anos) até funcionou.*

		Idade			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10	111	13,0	13,0	13,0
	11	162	19,0	19,0	32,0
	12	173	20,3	20,3	52,3
	13	150	17,6	17,6	70,0
	14	162	19,0	19,0	89,0
	15	71	8,3	8,3	97,3
	16	19	2,2	2,2	99,5
	17	3	,4	,4	99,9
	18	1	,1	,1	100,0
Total		852	100,0	100,0	

# Mas

- Se se fizer: **Analyze / Descriptive Statistics / Frequencies**
- Variáveis: **V1.1. (Idade)**
- Mas agora escolher o botão à direita **Statistics** posso ativar várias opções disponíveis à minha escolha (Percentis, medidas de tendência central, medidas de dispersão ou distribuição).
- **Continue** e **OK**.

Isto é só em variáveis quantitativas, ou seja as estatísticas só podem ser determinadas em variáveis QUANTITATIVAS.

**Frequencies: Statistics**

**Percentile Values**

☒ Quartiles

☐ Cut points for: 10 equal groups

☐ Percentile(s):

Add

Change

Remove

**Central Tendency**

☒ Mean

☒ Median

☐ Mode

☐ Sum

☐ Values are group midpoints

**Dispersion**

☒ Std. deviation

☐ Variance

☐ Range

☒ Minimum

☒ Maximum

☐ S.E. mean

**Characterize Posterior Dis...**

☐ Skewness

☐ Kurtosis

Continue Cancel Help

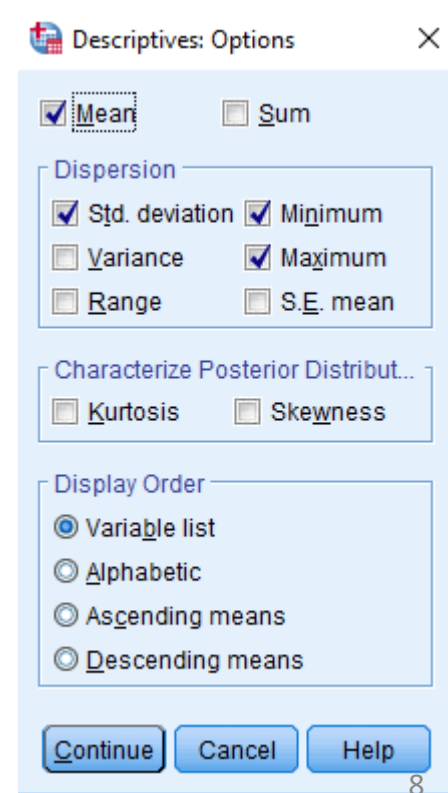
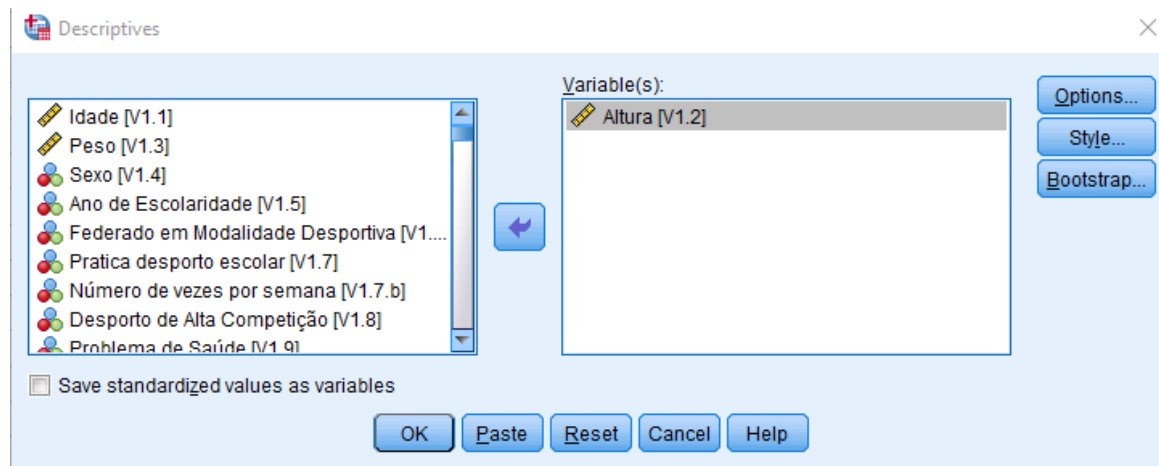
- Obtenho mais um quadro com as estatísticas pedidas:

## Statistics

Idade		
N	Valid	852
	Missing	0
Mean		12,47
Median		12,00
Std. Deviation		1,633
Minimum		10
Maximum		18
Percentiles	25	11,00
	50	12,00
	75	14,00

## 2. DESCRITIVOS

- DESCRIPTIVES – só para análise univariada em variáveis quantitativas
- Fazer. **Analyze / Descriptive Statistics / Descriptives**
- Variáveis: **V1.2 (Altura)**
- Escolher o botão à direita **Options** posso ativar várias opções disponíveis à minha escolha (média, soma, dispersão, ...):
- **Continue**
- **OK.**




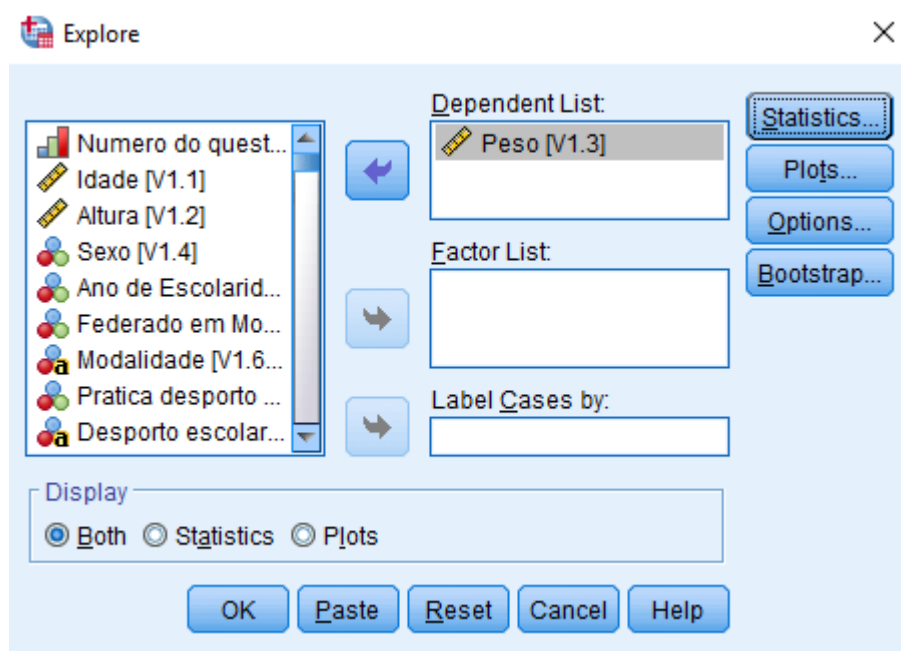


- Ver os **OUTPUTS**: aparece uma tabela mais sobre a horizontal com o que foi pedido.

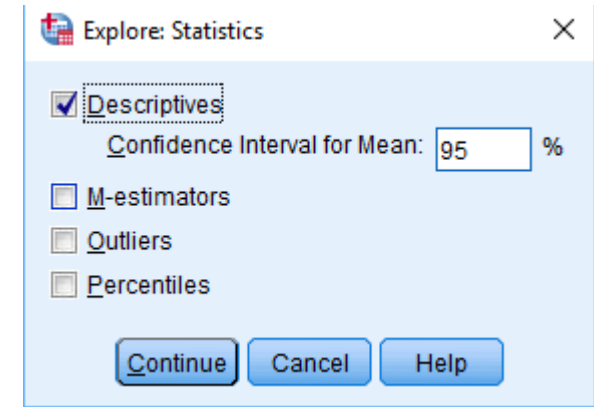
Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Altura	779	1,00	1,85	1,5749	,10964
Valid N (listwise)	779				

### 3. EXPLORAR

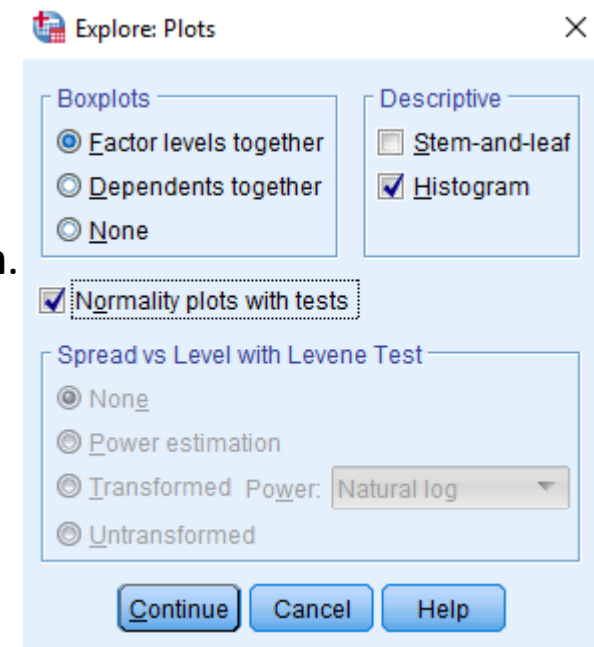
- EXPLORE – só em variáveis quantitativas
- Fazer: **Analyze / Descriptive Statistics / Explore**
- Variáveis: **V1.3 (Peso)**  vai para dependent list
- Por baixo aparece uma caixa (**Display**) que dá hipótese de escolher se quero obter só os gráficos, só as tabelas ou ambos.
- Deixo Both (opção por defeito)



- Escolher o botão à direita **Statistics**
- Aparece por defeito a opção: **Descriptives**
- Neste caso deixo assim e não escolho mais nada.
- Faço **Continue**.



- Escolher o botão à direita **Plots**
- Aparece por defeito a opção: Stem-and-leaf
- Neste caso desativo a que está e escolho: **Histogram**.
- Ativo também a opção: **normality plots with tests**.
- Faço **Continue** e **OK**.



- No OUTPUT aparecem 3 tabelas e 4 gráficos
- Tabela das estatísticas - descritivos:

### Descriptives

		Statistic	Std. Error
Peso	Mean	47,8595	,41391
	95% Confidence Interval for Mean	Lower Bound	47,0470
		Upper Bound	48,6720
	5% Trimmed Mean	47,3583	
	Median	47,0000	
	Variance	134,316	
	Std. Deviation	11,58947	
	Minimum	25,00	
	Maximum	98,00	
	Range	73,00	
	Interquartile Range	16,00	
	Skewness	,616	,087
	Kurtosis	,349	,174

### NOTA IMPORTANTE

Nos Outputs do SPSS para a interpretação da distribuição normal divide-se

Kurtosis (K)  
erro respstivo

Skewiness (SK)  
erro respstivo

Neste caso:

Se  $|K| > 2 \rightarrow$  desvio da normalidade

*Kurtosis:*  $0,349/0,174 = 2,00$

Se  $|SK| > 2 \rightarrow$  desvio da normalidade

*Skewiness:*  $0,616/0,087 = 7,08$

- Para relembrar a interpretação dos desvios à normalidade

#### CURTÓSE (K)



Curva platicurica

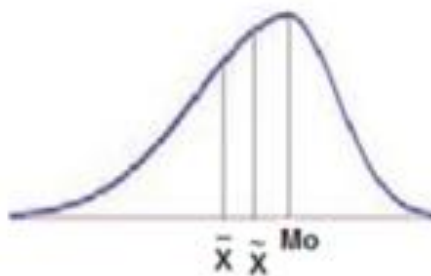
$$K < -2$$



Curva leptocurtica

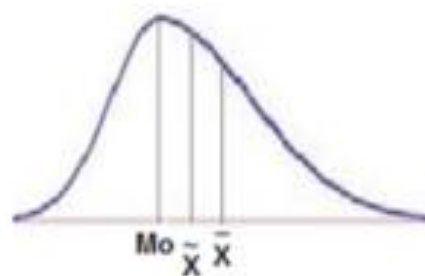
$$K > +2$$

#### ENVIESAMENTO (SK)



Curva enviesada à direita  
Skewness negativo

$$SK < -2$$



Curva enviesada à esquerda  
Skewness positivo

$$SK > +2$$

Neste caso:

Kurtosis:  $0,349/0,174 = 2,00$  ➡ **não** é platicurica nem leptocurtica

Skewness:  $0,616/0,087 = 7,08$  ➡ **é enviesada à esquerda**

- Tabela dos testes de Normalidade

**Tests of Normality**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Peso	,069	784	,000	,973	784	,000

a. Lilliefors Significance Correction

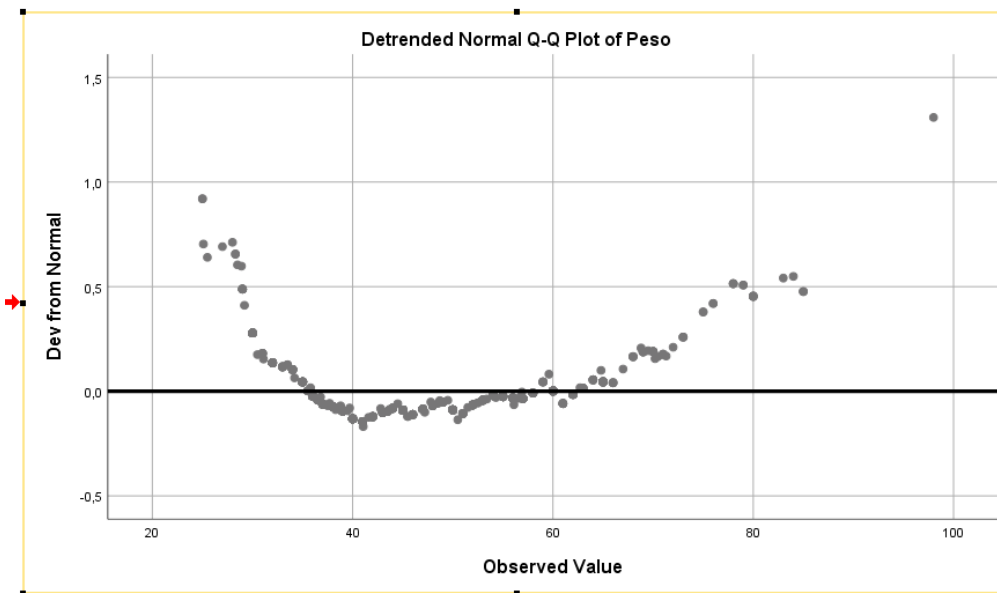
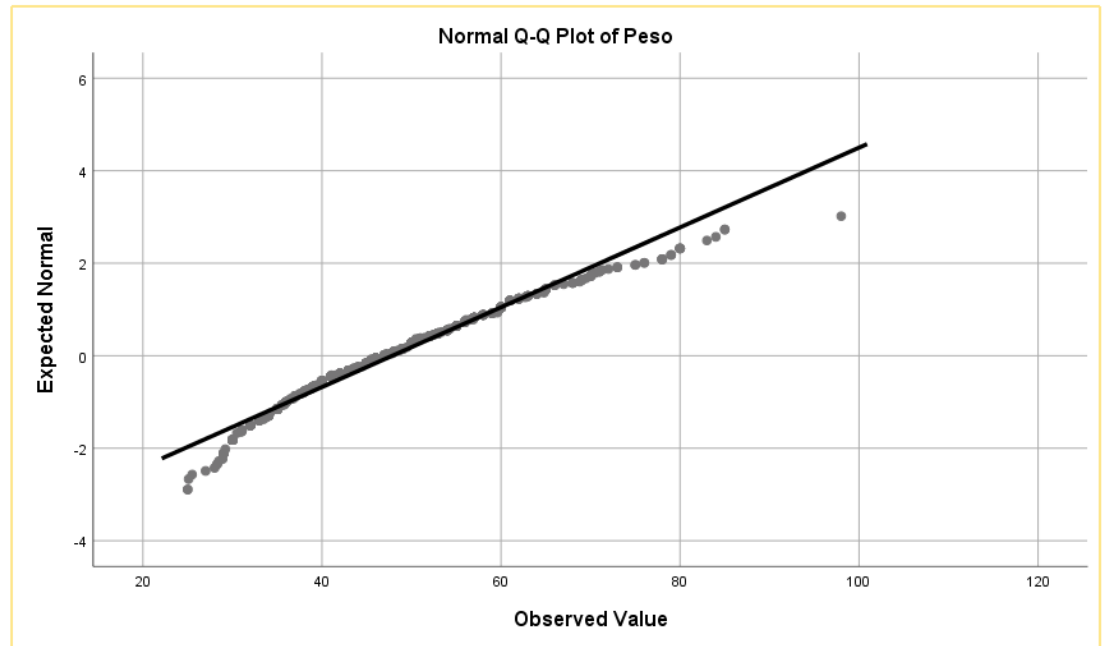
### Teste da normalidade (p)

Se  $N > 30 \Rightarrow$  Kolmogoronov – Smirnov

Se  $N < 30 \Rightarrow$  Shapiro – Wilk

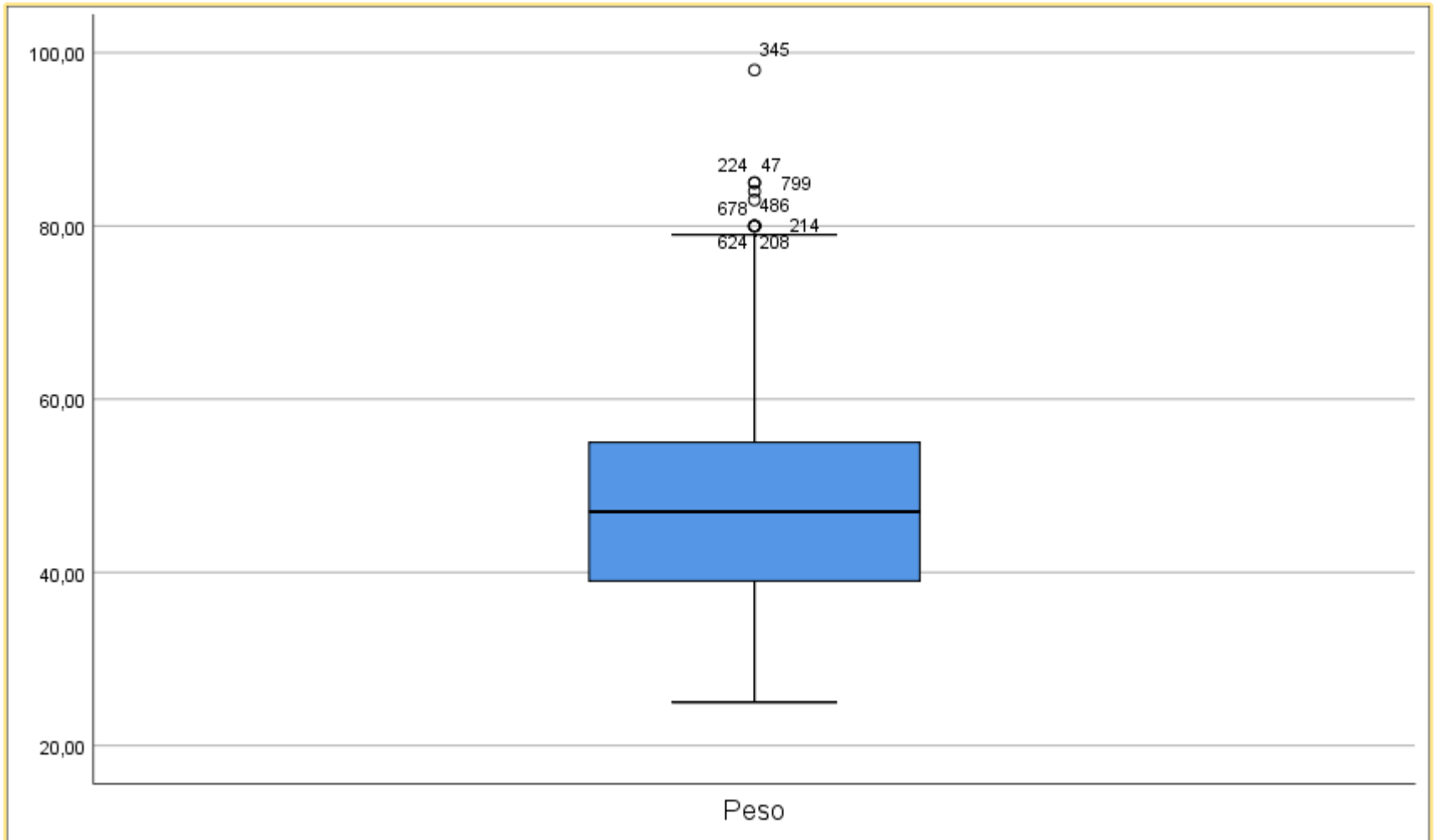
Se  $p(KS) < 5\% \Rightarrow$  há diferenças para a normalidade  
Curva não normal

## ■ Gráficos da normalidade



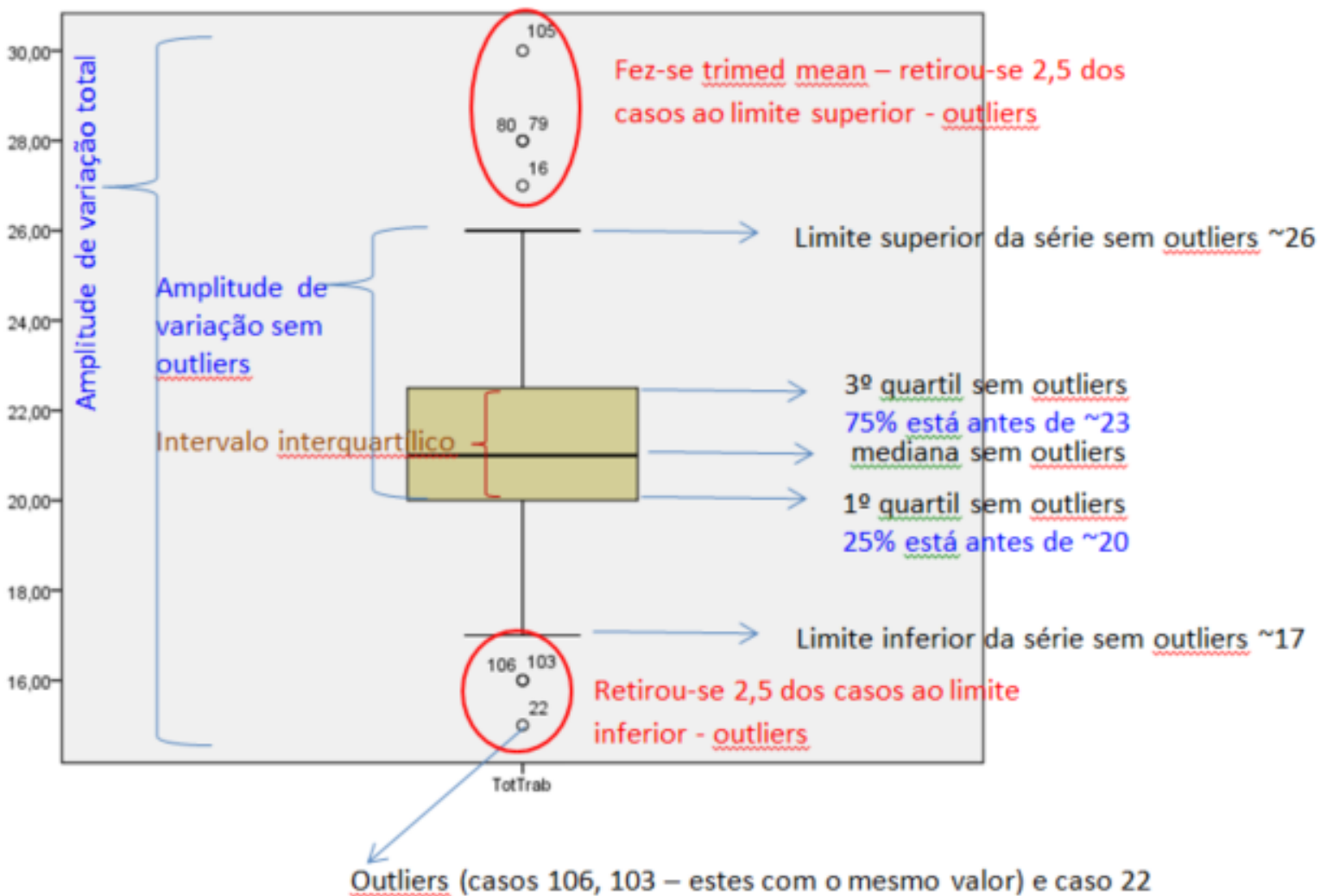
- ✓ Do gráfico da normalidade vê-se que não é uma distribuição normal, pois os pontos não ficam sobre a diagonal.
- ✓ Do gráfico de baixo verifica-se os erros

- Obtenho ainda o gráfico: Boxplot (Diagrama de Bigodes)
- A 5% trimmed mean (média aparada) retira os outliers – 5% (acima e/ou abaixo)

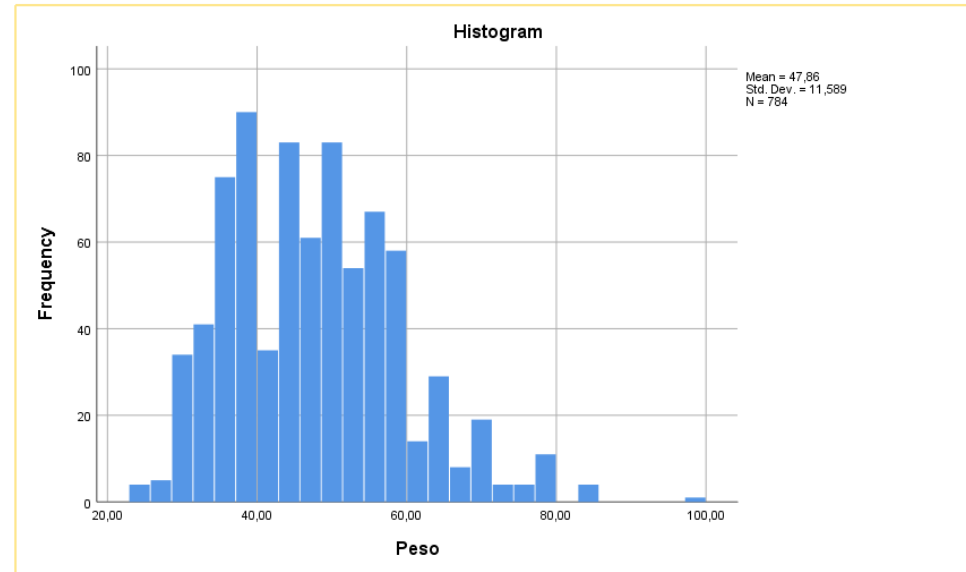




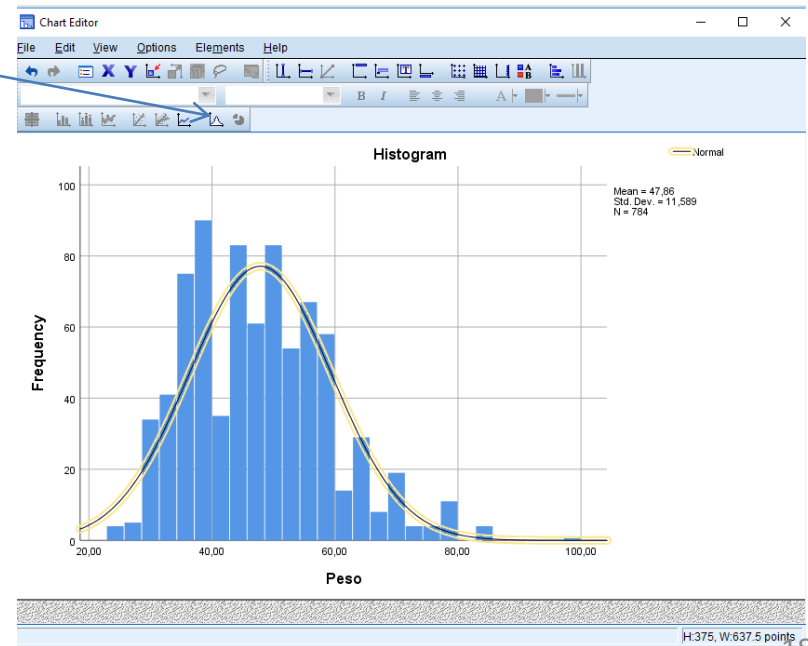
- Para relembrar a interpretação do Boxplot





- Obtenho também o histograma
- Clicamos duas vezes com o rato e abre uma janela com as propriedades (não fazemos nada e deixamos aberta).
- Clicamos no ICON (curva de gauss)



- Aparece a curva no histograma:
- É enviesada à esquerda (inicia-se acima do zero para o peso nulo).
- Para sair do modo editar gráfico faço close, nas propriedades e close na outra janela (no X do canto superior direito).



- Se quisesse saber as características da variável peso, mas diferenciando por sexo?
- Fazer: **Analyze / Descriptive Statistics / Explore**
- Variáveis: **V1.3 (Peso)**  vai para dependent list  
**V1.4 (Sexo)**  vai para factor list

*Neste caso, como eu já tinha feito antes a minha escolha de opções em **Statistics** e em **Plots**, elas ainda lá estão memorizadas.*

Só faço agora **OK**.


### Como reportamos os dados do **OUTPUT** para uma tabela

- A partir do output temos de criar uma tabela e interpretar os resultados.
- Como se faz uma tabela de estatísticas!

**Tabela 3.** Estatísticas relativas ao peso por sexos.

Sexo	N	Min	Max	$\bar{X}$	Desv.pad	CV	SK/ erro	K/ erro	p teste Norm.
Feminino	385	25,00	85,00	46,11	10,15	22,01%	4,19	1,85	0,001
Masculino	399	25,00	98,00	49,54	12,61	25,45%	4,43	-0,04	0,000
Totais	784	25,00	98,00	47,86	11,59	24,22%	7,08	2,00	0,000

- Para preencher os totais vai-se às estatísticas que se fazem sem dividir por sexos.
- Para calcular **CV** = desvio/Média x 100%
- Em todos os casos se vê o teste de normalidade pelo Kolmogorov-Smirnov porque  $N > 30$
- Neste caso, vemos ver a dimensão dos grupos:


 Como  $N_{\text{maior}}/n_{\text{menor}} = 399/385 = 1,04 < 1,5 \rightarrow$  os grupos são homogêneos  
 Podemos usar técnicas de análise paramétricas: como são dois grupos será o teste t de student

## Case Processing Summary

		Valid		Cases Missing		Total	
Sexo		N	Percent	N	Percent	N	Percent
Peso	Feminino	385	90,4%	41	9,6%	426	100,0%
	Masculino	399	93,9%	26	6,1%	425	100,0%

## Descriptives

			Statistic	Std. Error
Peso	Feminino	Mean	46,1145	,51730
		95% Confidence Interval for Mean	Lower Bound	45,0974
			Upper Bound	47,1316
		5% Trimmed Mean	45,7695	
		Median	45,0000	
		Variance	103,027	
		Std. Deviation	10,15024	
		Minimum	25,00	
		Maximum	85,00	
		Range	60,00	
		Interquartile Range	14,50	
		Skewness	,520	,124
		Kurtosis	,458	,248
	Masculino	Mean	49,5432	,63137
		95% Confidence Interval for Mean	Lower Bound	48,3020
			Upper Bound	50,7845
		5% Trimmed Mean	49,0469	
		Median	49,0000	
		Variance	159,054	
		Std. Deviation	12,61165	
		Minimum	25,00	
		Maximum	98,00	
		Range	73,00	
		Interquartile Range	19,00	
		Skewness	,540	,122
		Kurtosis	-,009	21,244

## Tests of Normality

		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
Sexo		Statistic	df	Sig.	Statistic	df	Sig.
Peso	Feminino	,063	385	,001	,979	385	,000
	Masculino	,084	399	,000	,973	399	,000

a. Lilliefors Significance Correction

**Tabela 3.**  
Estatísticas  
relativas ao peso  
por sexos.

Sexo	N	Min	Max	$\bar{X}$	Desv.pad	CV	SK/ erro	K/ erro	p teste Norm.
Feminino	385	25,00	85,00	46,11	10,15	22,01%	4,19	1,85	0,001
Masculino	399	25,00	98,00	49,54	12,61	25,45%	4,43	-0,04	0,000
Totais	784	25,00	98,00	47,86	11,59	24,22%	7,08	2,00	0,000

- Pelos coeficientes de variação notamos que existe uma dispersão moderada, porque os valores estão entre 15% e 30%.
- Os valores de Kurtosis revelam curvas sem desvios (porque os valores de K estão entre -2 e +2).
- Os valores de assimetria ou enviesamento revelam curvas enviesadas à esquerda (porque os valores de SK são superiores a 2).
- O teste de aderência à normalidade de Komogorov-Smirnov (N > 30 em todos os casos) revela que as distribuições não são normais (porque o valor de p é inferior a 0,05 e só seria normal se fosse > 5%).

Relembrar os valores de referência para a significância:

$P > 5\%$   $\Rightarrow$  **NÃO há diferenças** significativas

$P < 5\%$   $\Rightarrow$  Há diferenças **significativas**

$P < 1\%$   $\Rightarrow$  Há diferenças **bastante significativas**

$P < 0,1\%$   $\Rightarrow$  Há diferenças **altamente significativas**

- A variável **V1.3 (peso)** é quantitativa.
- Posso fazer a classificação – ou seja, criar grupos de corte



Permitem a classificação de uma variável quantitativa, ou seja, a transformação de uma var. quantitativa em qualitativa.

- Há duas formas que podem ser utilizadas para fazer esta operação:
  - 1) A primeira contempla a média e o desvio padrão:  $\bar{X}$  e  $s$
  - 2) A segunda contempla a mediana e o intervalo interquartílico:  $\tilde{X}$  e I.I.Q.



- Estabelecer grupos de corte com base na média e o desvio padrão:  $\bar{X}$  e  $s$

*Quando a amostra tem uma distribuição normal ou os enviesamentos/achatamentos são muito pequenos (p.ex. se  $SK$  até 7 e  $K$  até 3).*

Usa-se:  $\bar{X} \pm 0,25 s$

- Estabelecer grupos de corte com base na mediana e o intervalo interquartílico:

$\tilde{X}$  e I.I.Q.

*Se a distribuição não é normal, ou seja, se há grandes enviesamentos ou achatamentos (p.ex. se  $SK > 7$  e  $K > 3$ ).*

Usa-se:  $\tilde{X} \pm 0,25 I.I.Q.$

- No nosso caso, como os valores de desvio à normalidade estão até aos limites definidos vamos usar a 1ª opção:

SK/ erro	K/ erro
4,43	-0,04
4,19	1,85
7,08	2,00

1. Estabelecer grupos e corte para a amostra global
2. Estabelecer grupos e corte para o sexo feminino
3. Estabelecer grupos e corte para o sexo masculino

## 1. Estabelecer grupos e corte para a AMOSTRA GLOBAL

$$\bar{X} = 47,86 \quad ; \quad s = 11,59 \Rightarrow s/4 \text{ (ou } 0,25s) = 2,90$$

Temos então:  $47,86 \pm 2,90$

1º ponto de corte:  $47,86 - 2,90 = 44,96 \text{ kg}$

2º ponto de corte:  $47,86 + 2,90 = 50,76 \text{ kg}$

Assim,

1º grupo: peso até 44,96 kg ( $< 44,96 \text{ kg}$ )

2º grupo: peso entre 44,96 e 50,76 kg

3º grupo: peso superior ou igual a 50,76 kg ( $\geq 50,76 \text{ kg}$ )

## 2. Estabelecer grupos e corte para o SEXO FEMININO

$$\bar{X} = 46,11 \quad ; \quad s = 10,15 \Rightarrow s/4 \text{ (ou } 0,25s) = 2,54$$

Temos então:  $46,11 \pm 2,54 \Rightarrow$  pontos de corte: 43,57 kg e 48,65 kg

Assim,

1º grupo: peso até 43,57 kg

2º grupo: peso entre 43,57 e 48,65kg

3º grupo: peso superior ou igual a 48,65 kg

## 3. Estabelecer grupos e corte para o SEXO MASCULINO

$$\bar{X} = 49,54 \quad ; \quad s = 12,61 \Rightarrow s/4 \text{ (ou } 0,25s) = 3,15$$


Temos então:  $49,54 \pm 3,15 \Rightarrow$  pontos de corte: 46,39 kg e 52,69 kg

Assim,

1º grupo: peso até 46,39 kg

2º grupo: peso entre 46,39 e 52,69 kg

3º grupo: peso superior ou igual a 52,69 kg

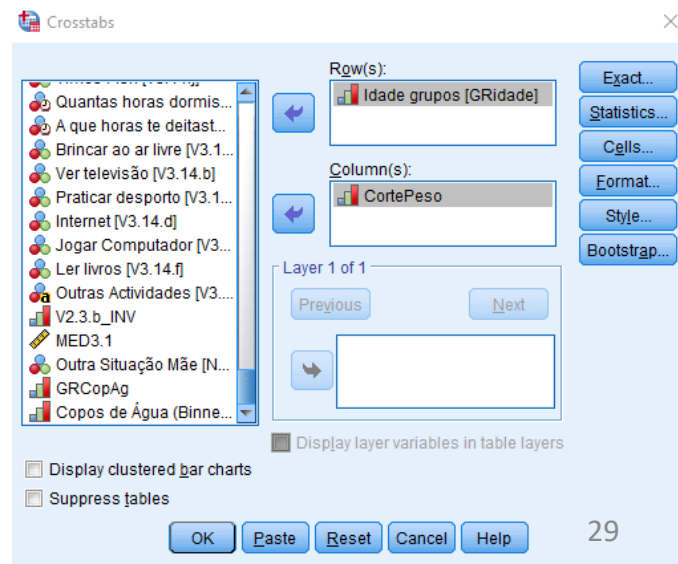
- Estabelecidos os grupos de corte, vamos criá-los na base de dados:  
exemplificação para o TOTAL DA MOSTRA:
  - Fazer: **Transform / Recode into Different Variable**
  - **V1.3 (Peso)**  Input Variable  $\Rightarrow$  Output Variable
  - Output variable Name: **CortePeso** **Change**
  - Old and New Values
    - LOWEST: 44,96  $\Rightarrow$  value 1 **Add**
    - Range: 44,96 through 50,76  $\Rightarrow$  value 2 **Add**
    - HIGHEST: 50,76  $\Rightarrow$  value 3 **Add**
  - Fazer **Continue** e **OK**.
  
- Depois, ir à vista **Variable View** para tirar as casas decimais, colocar variável como ordinal e especificar os valores.
- Se por exemplo quiséssemos fazer isto apenas para o sexo Masculino, primeiro íamos seleccionar só os homens: **Data / Select Cases** e depois era tudo igual.

## 4. TABELAS DE CONTINGÊNCIA (*CROSS TABS*) E TESTE DO $\chi^2$

- Comando **Descriptive Statistics / Crosstabs**: permite cruzar duas variáveis – é o *Cross Tabs* e teste do Qui-Quadrado
- Usar o Crosstabs - faz análise bivariada (só em variáveis qualitativas)
- Vamos verificar as classes de peso em função dos grupos de idade (atenção que estas duas variáveis originalmente eram de razão, mas nós vamos utilizar já a classificação em grupos em ambos os casos)
- Fazer: **Analyze / Descriptive Statistics / Crosstabs**
- Uma das variáveis vai para **Rows** e a outra vai para **Columns**

*NOTA: É indiferente colocar em linha ou em coluna a 1ª variável, mas por uma questão de estética coloca-se a variável que tem menos grupos em coluna.*

Idade = 4 grupos  $\Rightarrow$  Linhas  
Peso = 3 grupos  $\Rightarrow$  Colunas



- No botão **Statistics** à esquerda ativo a opção ☒ Chi-square
- Posso pedir os coeficientes de contingência ativando

☒ Contingency coefficient (na caixa Nominal)

*Nota: para tabelas 2x2 é o  $\Phi$  e para Tabelas de ordem superior é o coef. de contingência*

- Fazer **Continue**
- No botão **Cells** à esquerda por defeito está ☒ Observed
- Posso se quiser também pedir ☒ Expected (não o fiz neste caso pois não tem muita importância)
- Ativo as percentagens:
  - ☒ Row
  - ☒ Collumn
  - ☒ Total
- Ativo também ☒ Adjusted Standardized (na caixa dos residuals)
- Fazer **Continue** e **OK**.

- No **OUTPUT**:
- A 1ª tabela que aparece não tem muita importância

### Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
Idade grupos * CortePeso	784	92,0%	68	8,0%	852	100,0%

- A 2ª tabela é importante

### Idade grupos \* CortePeso Crosstabulation

			CortePeso			Total
			Peso < 44,96	peso no intervalo [44,96 50,67]	Peso >= 50,67	
Idade grupos	10 a 11 anos	Count	188	44	14	246
		% within Idade grupos	76,4%	17,9%	5,7%	100,0%
		% within CortePeso	58,6%	24,0%	5,0%	31,4%
		% of Total	24,0%	5,6%	1,8%	31,4%
		Adjusted Residual	13,7	-2,4	-11,9	
	12 a 13 anos	Count	115	80	100	295
		% within Idade grupos	39,0%	27,1%	33,9%	100,0%
		% within CortePeso	35,8%	43,7%	35,7%	37,6%
		% of Total	14,7%	10,2%	12,8%	37,6%
		Adjusted Residual	-,9	1,9	-,8	
	14 a 15 anos	Count	18	56	147	221
		% within Idade grupos	8,1%	25,2%	66,5%	100,0%
		% within CortePeso	5,6%	30,6%	52,5%	28,2%
		% of Total	2,3%	7,1%	18,8%	28,2%
		Adjusted Residual	-11,7	,8	11,3	
	16 a 18 anos	Count	0	3	19	22
		% within Idade grupos	0,0%	13,6%	86,4%	100,0%
		% within CortePeso	0,0%	1,6%	6,8%	2,8%
		% of Total	0,0%	0,4%	2,4%	2,8%
		Adjusted Residual	-4,0	-1,1	5,0	
Total		Count	321	183	280	784
		% within Idade grupos	40,9%	23,3%	35,7%	100,0%
		% within CortePeso	100,0%	100,0%	100,0%	100,0%
		% of Total	40,9%	23,3%	35,7%	100,0%

## Levantar a hipótese $H_0$ :

Não há diferenças significativas entre os pesos nos grupos de idade.

**Tabela 4.** Teste de Chi-quadrado para o estudo da relação entre o peso e a idade.

Peso	Baixo < 44,96		Médio [44,96;50,67[		Elevado ≥ 50,67		Total		Resíduos Ajustados		
Idade	N	%	N	%	N	%	N	%	PBaix	PMed	PElev
10 -11 anos	188	58,6	44	24,0	14	5,0	246	31,4	13,7	-2,4	-11,9
12-13 anos	115	35,8	80	43,7	100	35,7	295	37,6	-0,9	1,9	-0,8
14-15 anos	18	5,6	56	30,6	147	52,5	221	28,2	-11,7	0,8	11,3
16-18 anos	0	0,0	3	1,6	19	6,8	22	2,8	-4,0	-1,1	5,0
Total	321	100	183	100	280	100	784	100			
Teste $\chi^2$	$\chi^2 = 286,05$ ; p = 0,000 ; CC = 0,517										

- Para o preenchimento da Tabela 4, retiramos os valores nas linhas que dizem: % *within* CortePeso
- Neste caso avaliamos: para cada classe de peso como se distribui em termos de idades
- Podíamos fazer uma outra tabela, em que para cada classe de idade iríamos avaliar como se distribuía o peso (usávamos as % *within* Idade)



## Leitura da Tabela:

- Os alunos com peso baixo são majoritariamente com idades entre 10 e 11 anos (58,6%), não havendo nenhuns alunos com peso abaixo de 44,96 na classe de idades de 16 a 18 anos (Tabela 4).
- Na classe de peso médio a maior parte dos alunos são nas idades 12 a 13 anos (43,7), mas havendo ainda percentagens importantes nas classes de idades 10-11 anos (com 24,0%) e 14-15 anos (30,6%). Existe uma percentagem muito pequena de alunos com idades entre 16-8 anos com peso médio (1,6%).
- Na classe de peso elevado, a maioria dos alunos são do grupo de idades 12-13 anos.

*Nota: Nas tabelas basta falar nas % e não precisamos discutir os valores das frequências absolutas.*

## Notas sobre os RESÍDUOS AJUSTADOS:

- Os resíduos ajustados medem as diferenças entre as frequências observadas e as esperadas. Eles são altamente importantes porque o teste de  $\chi^2$  apenas nos diz se há ou não diferenças significativas (através do valor do p) mas numa visão global do problema, enquanto com os resíduos ajustados conseguimos verificar exatamente onde se localizam essas diferenças. Os resíduos ajustados são mais potentes que o teste de  $\chi^2$ .
- Quando os valores dos resíduos ajustados forem superiores a 2 (o valor exato é 1,96) as células onde os resíduos são superiores a 2 são aquelas onde se localizam as diferenças **(a castanho na Tabela 4)**. *De notar que neste caso só os valores positivos superiores a 2 é que contam (não é o valor absoluto, pelo que se der -3 não conta).*
- Só se podem usar os resíduos ajustados quando em linha ou em coluna houver pelo menos 8 casos nas frequências observadas – no nosso problema há 2 casos em que isso não acontece – ~~estão rasurados~~.

- Terceira tabela do **OUTPUT**:

*Desta tabela só interessa ler os valores de Pearson Chi-square e respectiva significância: usados para preencher a última linha da Tabela 4*

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	286,056 <sup>a</sup>	6	,000
Likelihood Ratio	326,299	6	,000
Linear-by-Linear Association	275,767	1	,000
N of Valid Cases	784		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,14.

- É muito importante a nota no final da tabela, que especifica que há 0% de células que têm frequências esperadas inferiores a 5.



Isto é uma condição fundamental para a aplicação do teste de  $\chi^2$ .

- Se a % fosse superior a 20% já não poderia aplicar o teste.



Nesse caso teria de aplicar o teste correspondente não paramétrico - **teste de Fisher**, mas que só dá para tabelas 2x2, e implicaria a recodificação das variáveis em casos de dimensão superior a 2x2.

## Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	286,056 <sup>a</sup>	6	,000
Likelihood Ratio	326,299	6	,000
Linear-by-Linear Association	275,767	1	,000
N of Valid Cases	784		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,14.

- ✓ Neste caso pelo valor do p do  $\chi^2$  ( $p = 0,000 < 0,05$ ) verifica-se que há diferenças altamente significativas ao nível do peso entre as classes de idades dos alunos.



Rejeito  $H_0$ : Não há diferenças significativas entre os pesos nos grupos de idade e aceito  $H_1$ : Há diferenças significativas entre os pesos nos grupos de idade

- ✓ Através dos resíduos ajustados (Tabela 4) podemos dizer que as diferenças se localizam entre as pessoas de peso baixo com idade 10-11 anos, de peso médio com idade 12-13 anos e de peso elevado com idade 14-15 anos.

## Symmetric Measures

- Quarta Tabela do **OUTPUT** :

		Value	Approximate Significance
Nominal by Nominal	Contingency Coefficient	,517	,000
N of Valid Cases		784	

- Já verificámos que a associação entre as variáveis estudadas: peso e Idade é significativa
- O coeficiente de contingência permite avaliar a força dessa associação entre as variáveis
- Gama de variação:  $0 \leq CC \leq 1$
- Interpretação:
  - 0 a 0.1  $\Rightarrow$  associação fraca ou nenhuma
  - 0.1 a 0.3  $\Rightarrow$  associação baixa
  - 0.3 a 0.5  $\Rightarrow$  associação moderada
  - $\geq 0.5$   $\Rightarrow$  associação forte
- No nosso caso  $CC = 0,517$ , pelo que se considera uma associação forte.

**FIM**