UNIVERSIDADE DO ALGARVE
ESCOLA SUPERIOR DE SAÚDE

Investigação Aplicada I

Aula 7

1º Semestre 2016/17

Licenciatura em Ciências Biomédicas Laboratoriais

igrodrigues@ualg.pt; ESSUAlg: gabinete 2.06

Prof. Inês Rodrigues

Inferência estatística

Inferir dados sobre uma população através de uma amostra dessa população

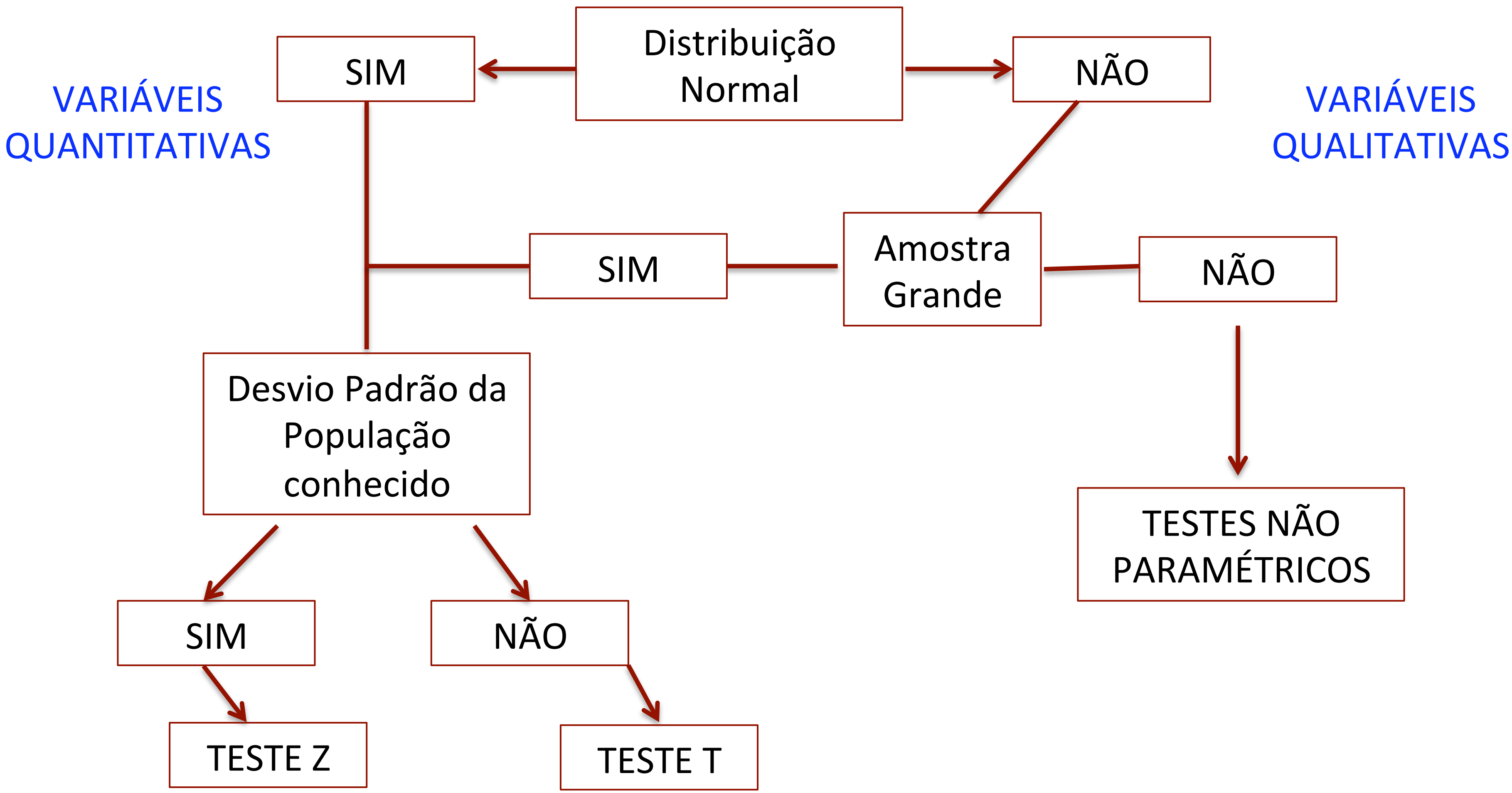
Desvio padrão da população é substituído pelo desvio padrão da amostra

Introdução de um erro no processo de inferência: erro de estimação do desvio padrão

Novo intervalo do desvio padrão será maior do que o considerado com a estatística Z

Distribuição t

Inferência estatística

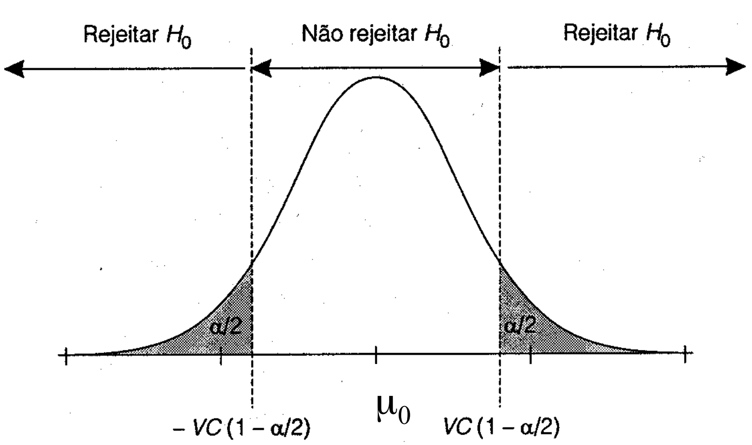


Inferência estatística

Inferir dados sobre uma população através de uma amostra dessa população

Desvio padrão da população é substituído pelo desvio padrão da amostra

Nome	Formula	Características e Notas
Estatística Z	$z = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$	(População segue uma distribuição normal **ou** $n > 30$) **e** σ (desvio-padrão) é conhecido (z é a distância da média em relação ao desvio padrão da média).
Estatística t	$t = \frac{\overline{x} - \mu_0}{(s/\sqrt{n})},$	(População segue uma distribuição normal **ou** $n > 30$) **e** σ desconhecido.



H_0: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$

Inferência estatística

Distribuição t

Tem em conta o desvio padrão amostral que varia em função dos graus de liberdade

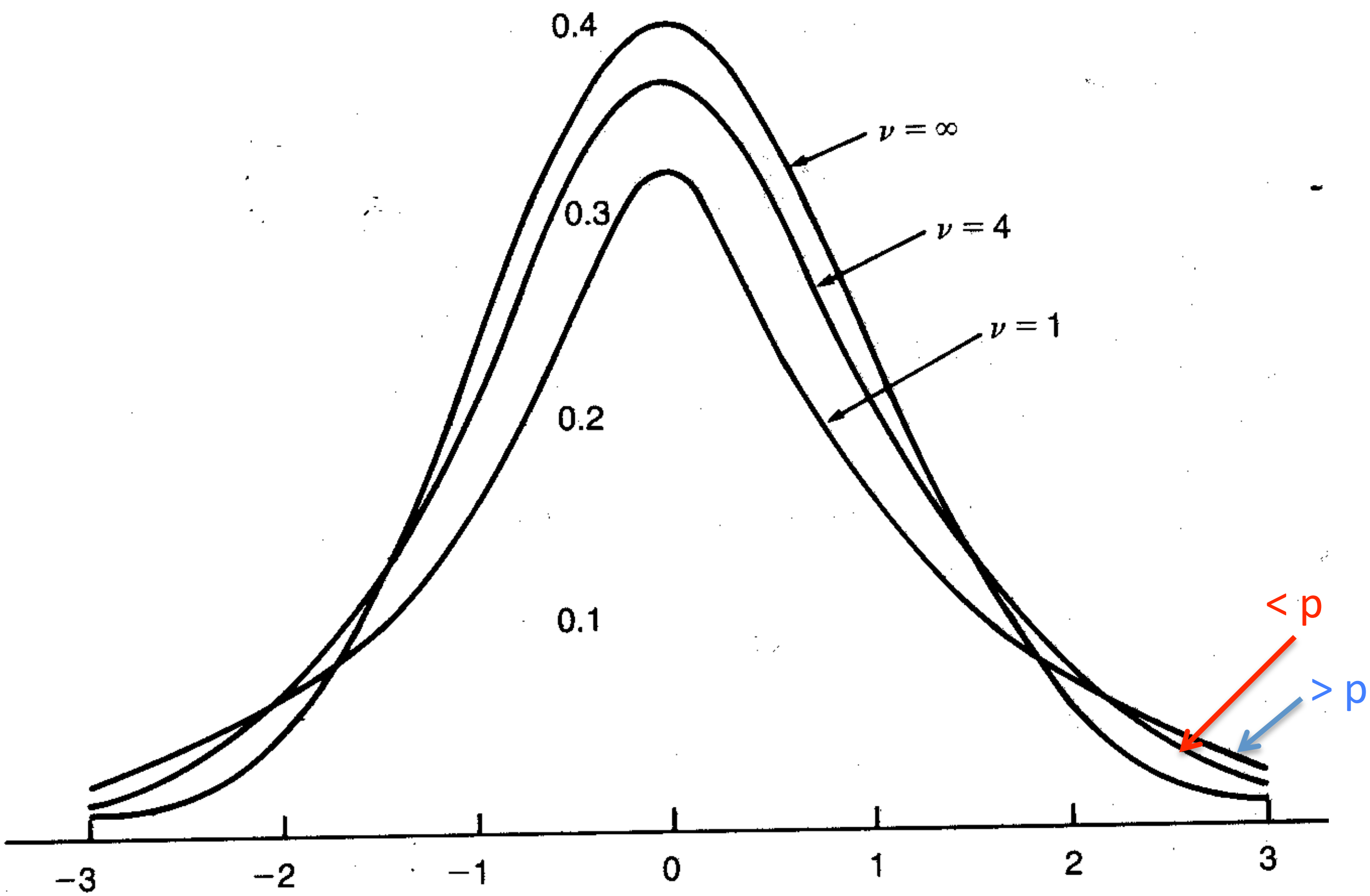
Parâmetro usado para descrever a distribuição t é o numero de graus de liberdade (*d.f. Degrees of freedom*), que consiste no tamanho da amostra menos o numero de amostras estimadas de uma população:

df = n-1 (uma amostra)

df = n-2 (para duas amostras emparelhadas)

...

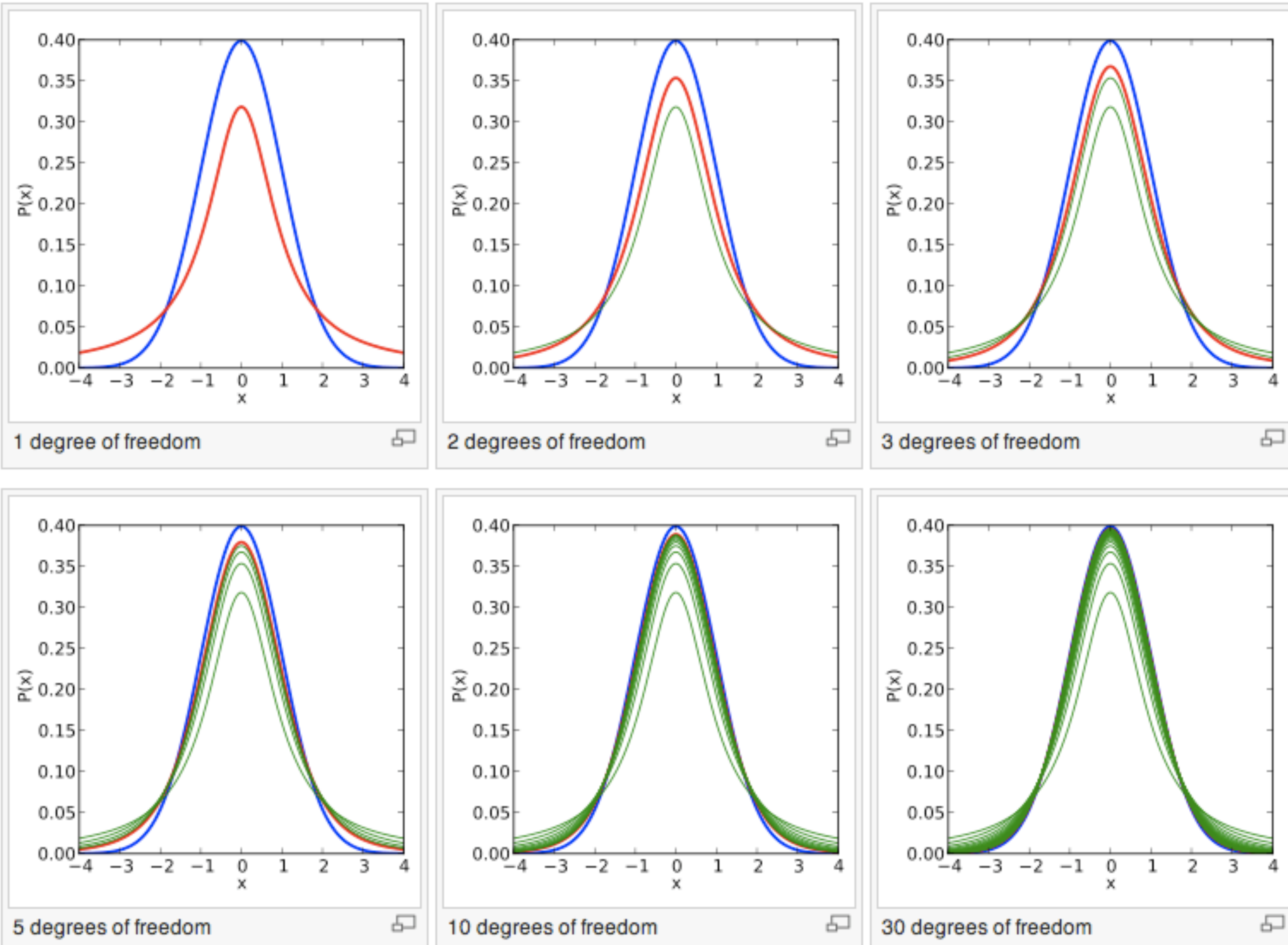
A distribuição t-Student



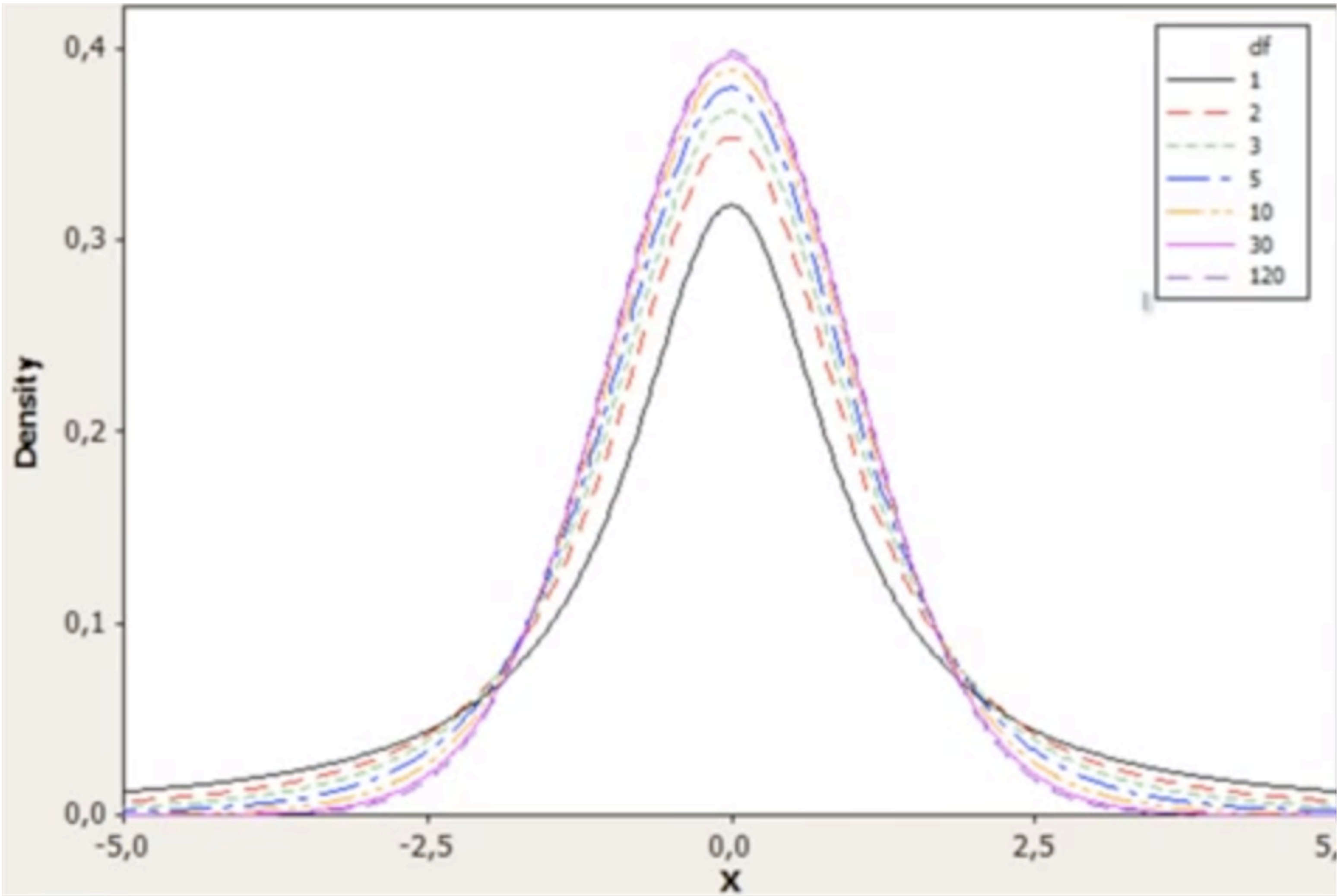
Representação gráfica de distribuições t reduzidas, para 1, 4 e ∞ graus de liberdade

A distribuição t-Student

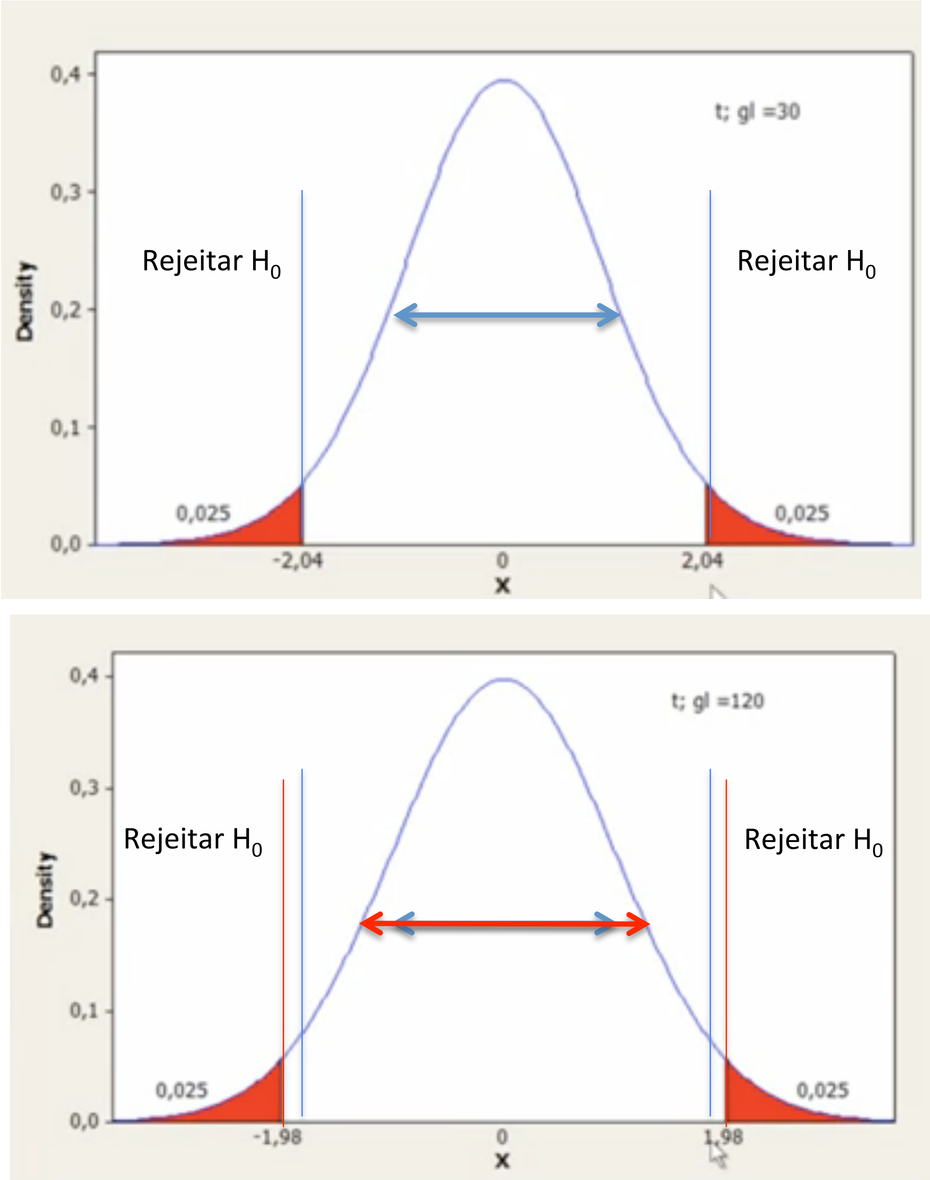
Distribuição Normal
Distribuição t



A distribuição t-Student



A distribuição t-Student



↔ gl = 30

↔ gl = 120

A distribuição t-Student

Para duas amostras existem dois desvios padrões da população que são desconhecidos

$$\boldsymbol{t} = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

Para várias amostras, vários desvios padrão

É necessário avaliar a variância das diferentes amostras: ANOVA

Análise simples de variância

- Analisa a variância das amostras. Define se são independentes ou de igual variabilidade (*i.e.* emparelhadas).

- A variabilidade total (SQ(T)), a variabilidade devida aos tratamentos (SQ(TR)) e a variabilidade residual (SQ(E)) são definidas pelas seguintes somas de quadrados:

$$\mathrm{SQ(T)} = \sum\sum\left(x_{ij} - \overline{\overline{X}}\right)^2 = \sum\sum x_{ij}^2 - \frac{\left(\sum\sum x_{ij}\right)^2}{n_t}$$

$$SQ(TR) = \sum n_j\left(\overline{x}_j - \overline{\overline{X}}\right)^2 = \sum_j \frac{\left(\sum_i x_{ij}\right)^2}{n_j} - \frac{\left(\sum\sum x_{ij}\right)^2}{n_t}$$

$$\mathrm{SQ(E)} = \sum\sum\left(x_{ij} - \overline{X}_j\right)^2 = \mathrm{SQ(T)} - \mathrm{SQ(TR)}$$

Análise simples de variância

- Decomposição dos graus de liberdade:
  - $\nu(T) = \nu(TR) + \nu(E)$ ou seja (nt –1) = (nt - r) + (r – 1)

Origem da variância	SQ	ν	QM
Diferenças “entre” grupos (TR)	$\sum n_j \left(\overline{X}_j - \overline{\overline{X}}\right)^2$	r - 1	$\dfrac{\sum n_j \left(\overline{X}_j - \overline{\overline{X}}\right)^2}{r-1}$
Diferenças “dentro” dos grupos (E)	$\sum\sum \left(x_{ij} - \overline{X}_j\right)^2$	nt - r	$\dfrac{\sum\sum \left(x_{ij} - \overline{X}_j\right)^2}{n_t - r}$
TOTAIS (T)	$\sum\sum \left(x_{ij} - \overline{\overline{X}}\right)^2$	nt - 1	

Análise simples de variância

- **Teste de diferença entre os grupos (<u>Teste F</u>)**
  - QM (E) – Diferenças nas repetições dentro do grupo. Permite apreciar factores não controláveis e acidentalmente atuantes que contribuem para a variância na sua totalidade
  - QM (TR) – Diferenças das médias dos grupos para a média global. Reflete a variância das médias obtidas para cada tratamento em relação à média global.
  - Se todas as μj (amostras de uma população) forem iguais, então QM (TR) = 0
- Comparação de médias da amostra **: ANOVA**

Análise simples de variância

- **Teste de diferença entre os grupos : avaliação da igualdade entre variancias**

- **Teste F ou Teste de Levene**

  - H_0: $\mu1= \mu2= \mu3= ...=\mu j$
  - H_1: Nem todas as μj são iguais

$$F^* = \frac{QM(TR)}{QM(E)}$$

Para comparação de médias da amostra : t teste ou ANOVA

Inferência estatística

Distribuição *t* para duas amostras independentes

Exemplo:

H_0 : O peso dos recém-nascidos com mães diabéticas é $\leq$ que o peso dos recém-nascidos de mães não diabéticas

H_a : O peso dos recém-nascidos com mães diabéticas é > que o peso dos recém-nascidos de mães não diabéticas

α = 0,05

n = 210
Teste f → p= 0,01

Teste t (amostras ID; variância diferentes)→ p < 0,01

Aceito H_a

Inferência estatística

Distribuição *t* para duas amostras Emparelhadas

Dados emparelhados são aqueles que são obtidos aos pares, ou seja, o individuo é o controlo de si mesmo.

A mesma variável é medida em dois tempos diferentes (i.e. antes e depois de um tratamento) ou em dois sítios anatómicos diferentes no mesmo individuo (i.e. testar um medicamento apenas num olho; testar pomadas diferentes em zonas diferentes da pele de um indivíduo).

Inferência estatística

Distribuição *t* para duas amostras Emparelhadas

Exemplo: Um estudo compara o efeito de uma pomada oftálmica com a simples higienização do olho em casos de conjuntivite em pacientes.
Olho direito: pomada
Olho esquerdo: antisséptico

Seja D_i a diferença (no mesmo individuo) entre as observações registadas, para n pares.
A média das diferenças de todos os indivíduos será $\underline{D}$, e o desvio padrão das diferenças S_D

$H_{0\,:}$ $\mathbf{U_D} = 0$

H_A : $\mathbf{U_D}$ diferente de 0

T = $\underline{D}$ / SEM_D e gl=n-1
→ valor de p

Onde $\mathbf{U_D}$ é a média populacional sobre a qual desejo inferir

Exercício exemplo: t Test

- Vinte doentes foram distribuídos aleatoriamente por dois grupos de tratamento (10 doentes em cada grupo). Os seguintes dados correspondem à alteração sofrida pelos iões cloro do soro, depois de cada tratamento:

Tratamento A: 4.3 6.2 4.4 8.2 0.5 2.6 4.2 4.1 5.6 3.4
Tratamento B: 6.1 0.9 0.7 0.8 1.3 3.1 1.9 3.9 2.1 0.1

O que pode concluir relativamente à eficácia dos tratamentos?

Exercício exemplo: t Test

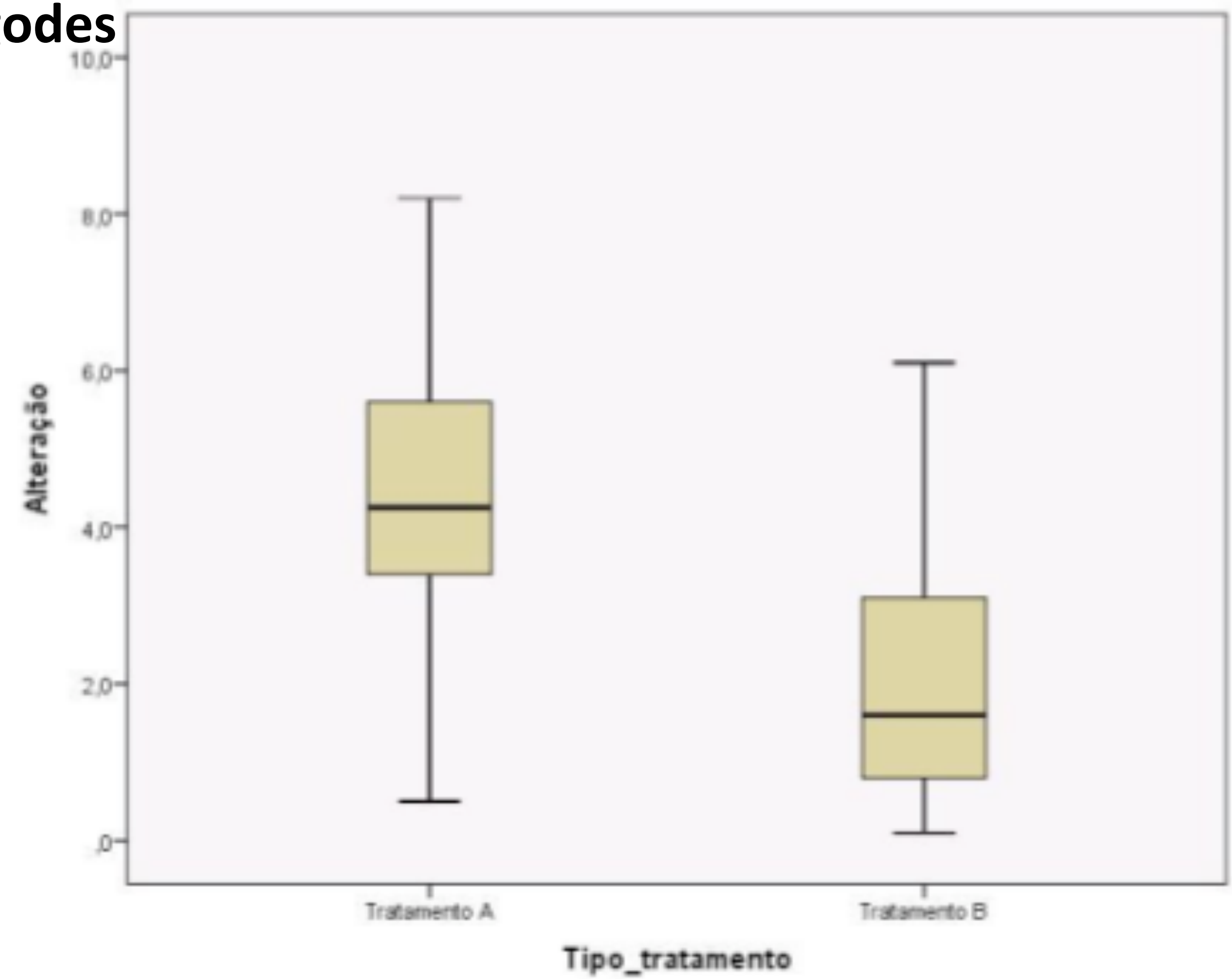
Descritivas

Descriptives

	Tipo_tratamento			Statistic	Std. Error
Alteração	1 Tratamento A	Mean		4,350	,6556
		95% Confidence Interval for Mean	Lower Bound	2,867	
			Upper Bound	5,833	
		5% Trimmed Mean		4,350	
		Median		4,250	
		Variance		4,298	
		Std. Deviation		2,0732	
		Minimum		,5	
		Maximum		8,2	
		Range		7,7	
		Interquartile Range		2,6	
		Skewness		,032	,687
		Kurtosis		1,139	1,334
	2 Tratamento B	Mean		2,090	,5774
		95% Confidence Interval for Mean	Lower Bound	,784	
			Upper Bound	3,396	
		5% Trimmed Mean		1,978	
		Median		1,600	
		Variance		3,334	
		Std. Deviation		1,8260	
		Minimum		,1	
		Maximum		6,1	
		Range		6,0	
		Interquartile Range		2,5	
		Skewness		1,290	,687
		Kurtosis		1,450	1,334

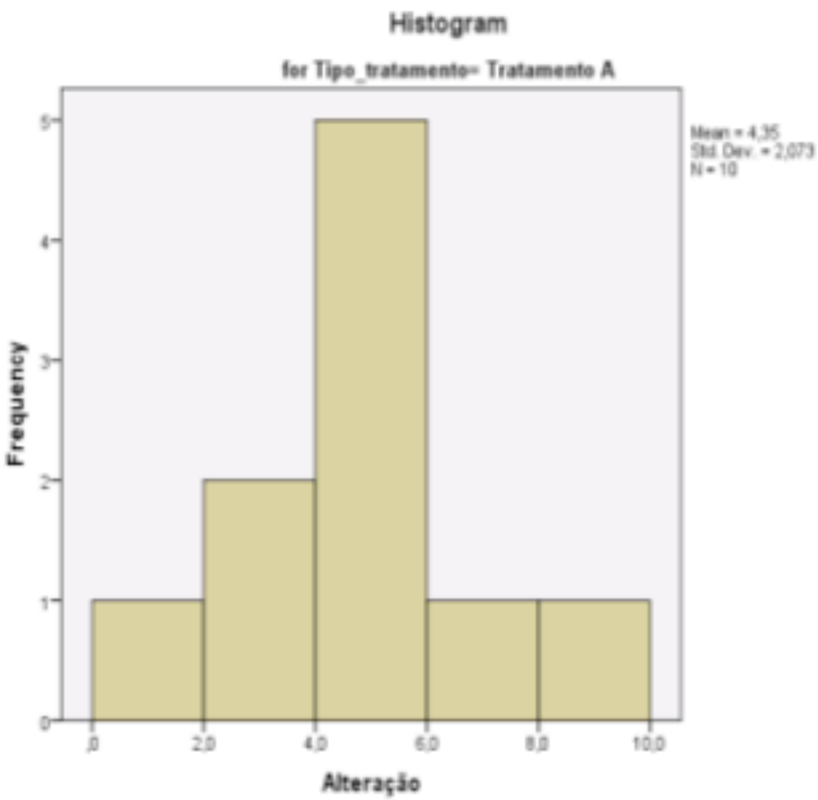
Exercício exemplo: t Test

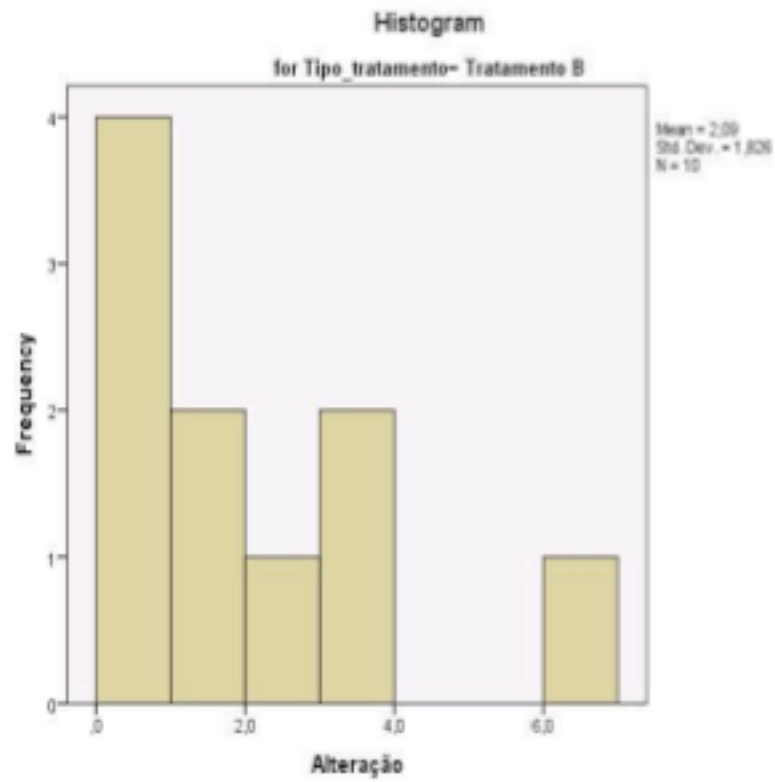
Caixas de Bigodes



Exercício exemplo: t Test

Histogramas





Exercício exemplo: t Test

Testes à normalidade

$$\begin{cases} H_0: & \text{Os dados do tratamento A/B seguem distribuição Normal} \\ H_1: & \text{Os dados do tratamento A/B não seguem distribuição Normal} \end{cases}$$

n<50

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Tipo tratamento	Statistic	df	Sig.	Statistic	df	Sig.
Alteração	1 Tratamento A	,190	10	,200*	,966	10	,856
	2 Tratamento B	,198	10	,200*	,887	10	,157

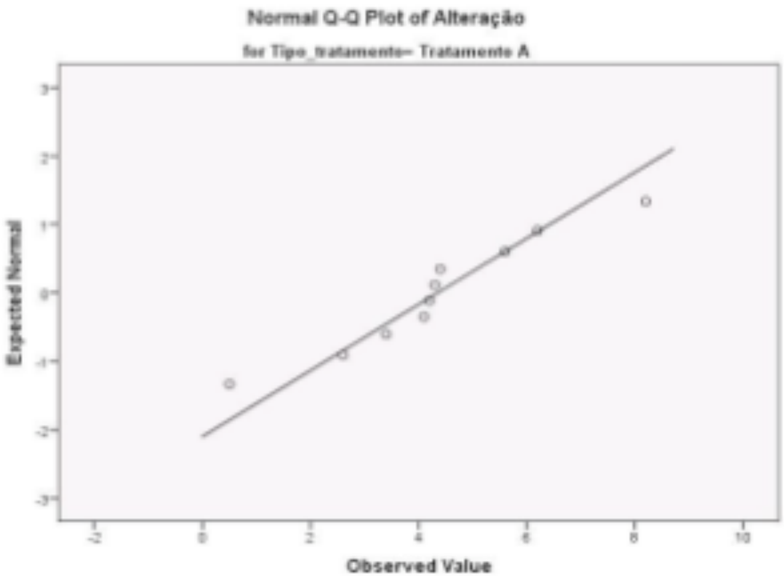
*. This is a lower bound of the true significance.

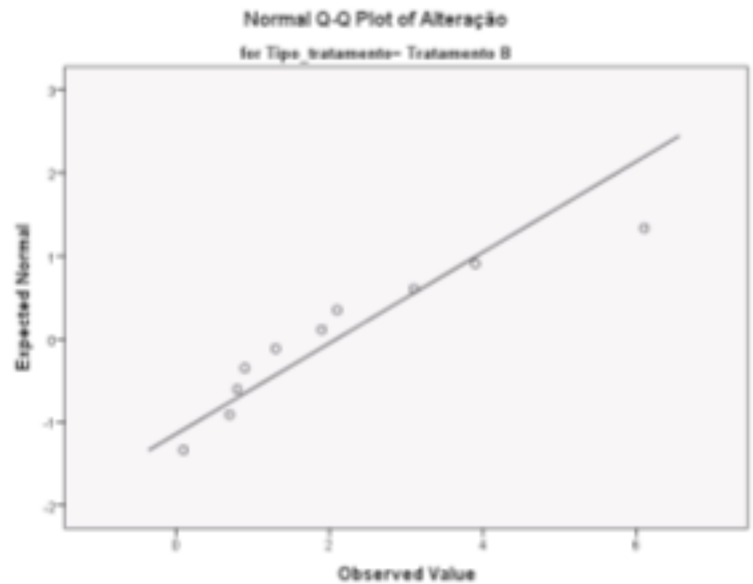
a. Lilliefors Significance Correction

Não se rejeita a hipótese de Normalidade

Exercício exemplo: t Test

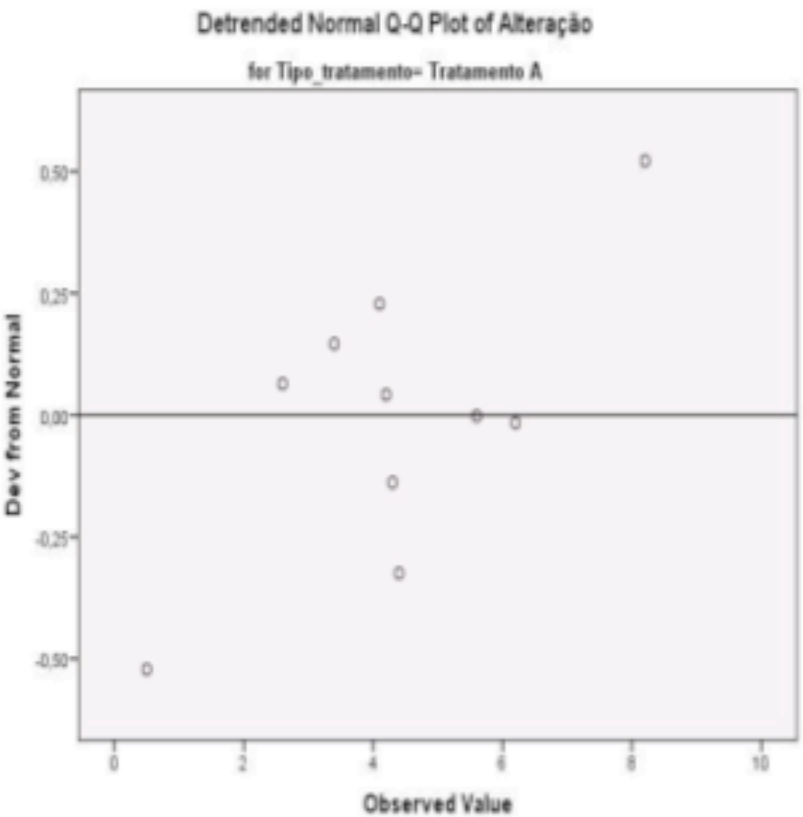
Testes à normalidade

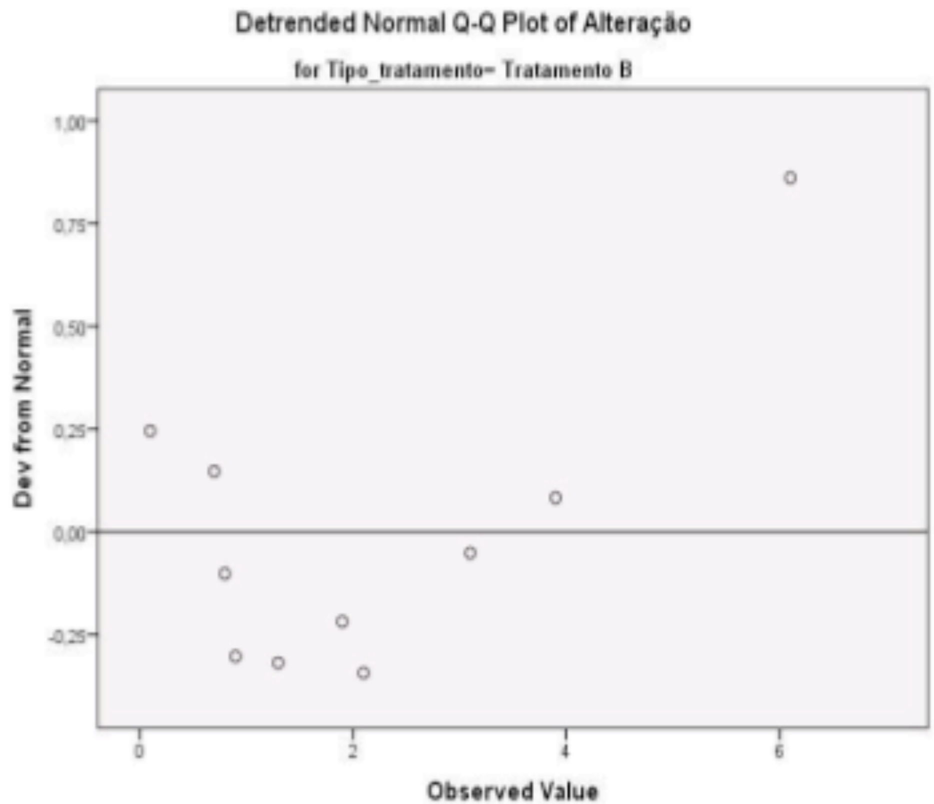




Exercício exemplo: t Test

Testes à normalidade





Exercício exemplo: t Test

T Test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
cloro	Equal variances assumed	,003	,957	2,587	18	,019	2,2600	,8737	,4245	4,0955
	Equal variances not assumed			2,587	17,717	,019	2,2600	,8737	,4224	4,0976

- A normalidade dos dados foi aceite.

- Os dados distribuem-se por uma amostra reduzida (n=10).
  O teste t deve ser confirmado por uma teste não paramétrico (Mann Whitney).

- As duas amostras são independentes.

Exercício exemplo: t Test

Teste não paramétrico corrobora os resultados do t Test

$$H_0: \chi_{0.5_A} - \chi_{0.5_B} = 0 \quad vs \quad H_1: \chi_{0.5_A} - \chi_{0.5_B} \neq 0$$

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Alteração is the same across categories of Tipo_tratamento.	Independent-Samples Mann-Whitney U Test	,019[1]	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

[1]Exact significance is displayed for this test.

Existe evidência de diferenças nas alterações dos iões entre os tratamentos A e B.

Amostra: 3 grupos

Exercício exemplo: Anova

Foram inquiridos 54 estudantes da ESSS de três turmas distintas, cada uma com 18 alunos. A cada aluno foi perguntada a sua classificação na disciplina de estatística e o seu nível de motivação, que poderá assumir uma de três categorias: baixa, média e elevada. O regente da disciplina procurou aferir se existiam diferenças significativas entre as médias atribuídas a cada uma das classificações e, para tal, recorreu à ANOVA. Os dados recolhidos encontram-se no quadro seguinte:

Baixa	14	12	9	15	15	10	11	11	10	14	12	11	15	12	12	14	13	12
Média	12	11	14	13	16	15	13	14	13	12	13	14	13	15	16	14	13	13
Elevada	17	16	16	18	16	17	14	15	16	12	18	13	18	14	16	17	15	17

O que pode concluir?

Exercício exemplo:
Anova

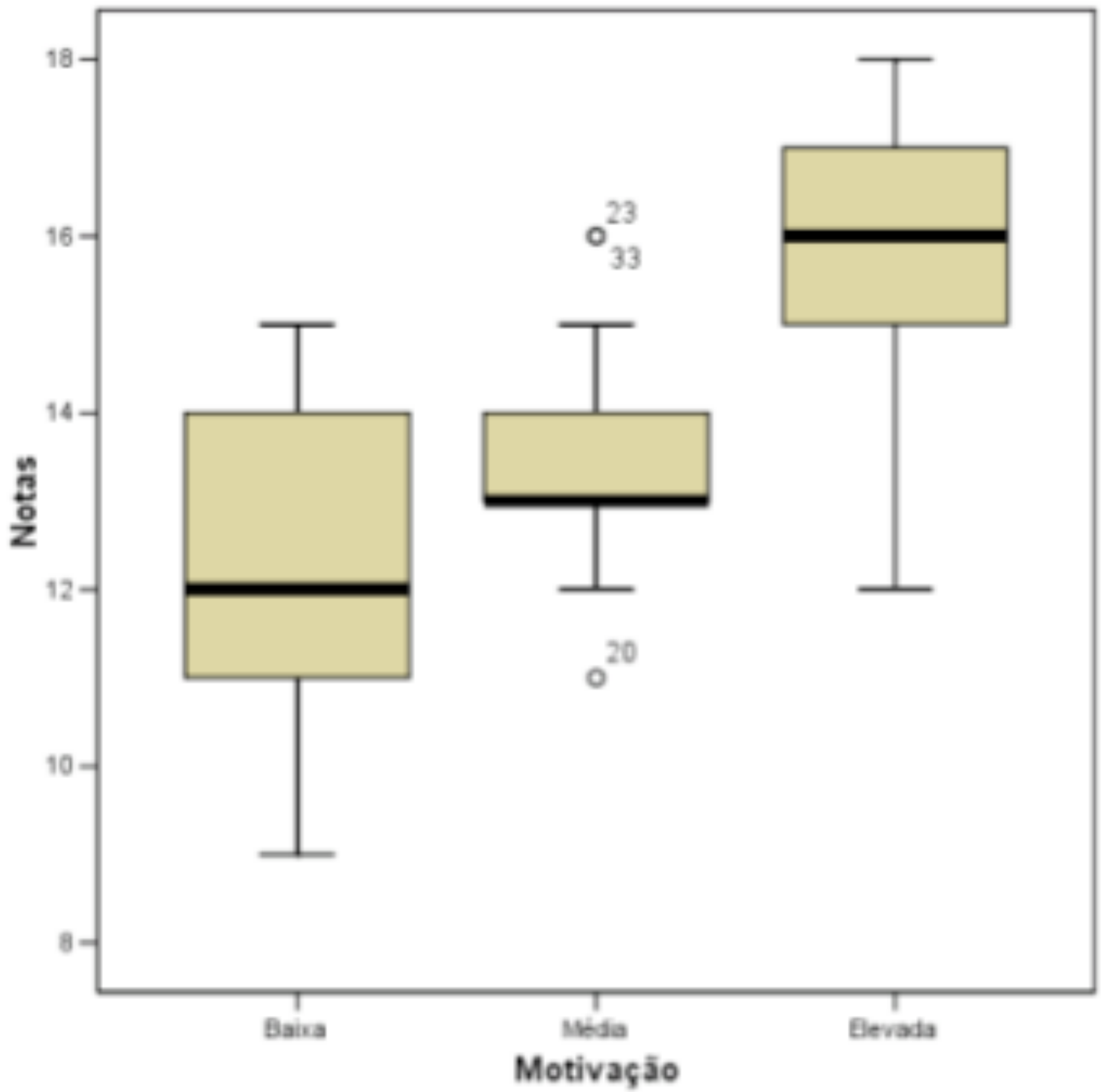
Descritivas

Descriptives

	Motivação			Statistic	Std. Error
Notas	1 Baixa	Mean		12,33	,435
		95% Confidence Interval for Mean	Lower Bound	11,41	
			Upper Bound	13,25	
		5% Trimmed Mean		12,37	
		Median		12,00	
		Variance		3,412	
		Std. Deviation		1,847	
		Minimum		9	
		Maximum		15	
		Range		6	
		Interquartile Range		3	
		Skewness		,014	,536
		Kurtosis		-,967	1,038
	2 Média	Mean		13,56	,315
		95% Confidence Interval for Mean	Lower Bound	12,89	
			Upper Bound	14,22	
		5% Trimmed Mean		13,56	
		Median		13,00	
		Variance		1,791	
		Std. Deviation		1,338	
		Minimum		11	
		Maximum		16	
		Range		5	
		Interquartile Range		1	
		Skewness		,281	,536
		Kurtosis		-,076	1,038
	3 Elevada	Mean		15,83	,406
		95% Confidence Interval for Mean	Lower Bound	14,98	
			Upper Bound	16,69	
		5% Trimmed Mean		15,93	
		Median		16,00	
		Variance		2,971	
		Std. Deviation		1,724	
		Minimum		12	
		Maximum		18	
		Range		6	
		Interquartile Range		2	
		Skewness		-,720	,536
		Kurtosis		-,027	1,038

Exercício exemplo: Anova

Caixas de Bigodes



Exercício exemplo: Anova

Normalidade

1) verificar a Normalidade para cada grupo/tratamento

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Motivação	Statistic	df	Sig.	Statistic	df	Sig.
Notas	1 Baixa	,183	18	,115	,934	18	,229
	2 Média	,217	18	,025	,932	18	,211
	3 Elevada	,205	18	,044	,923	18	,146

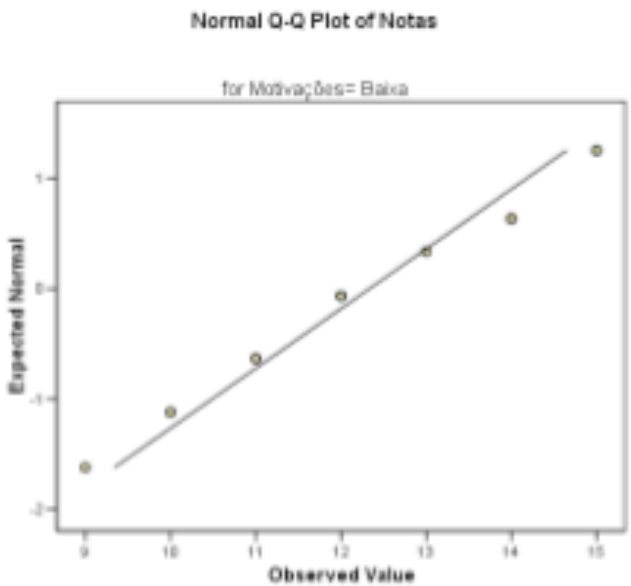
a. Lilliefors Significance Correction

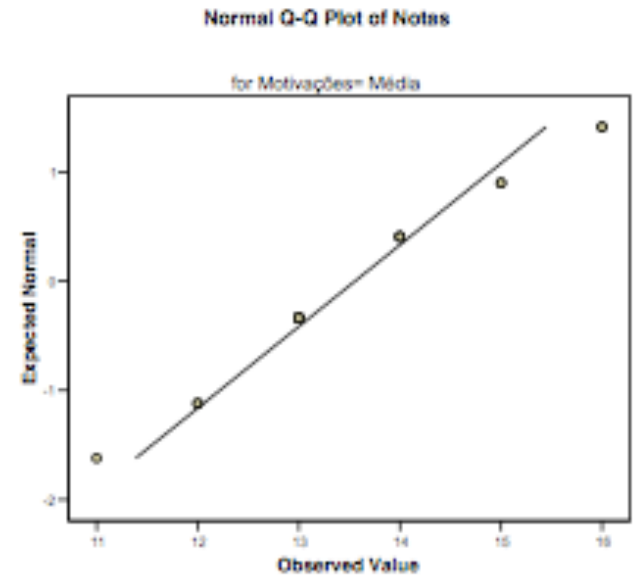
Não se rejeita H_0 ,
ou seja a normalidade
das notas da motivação.

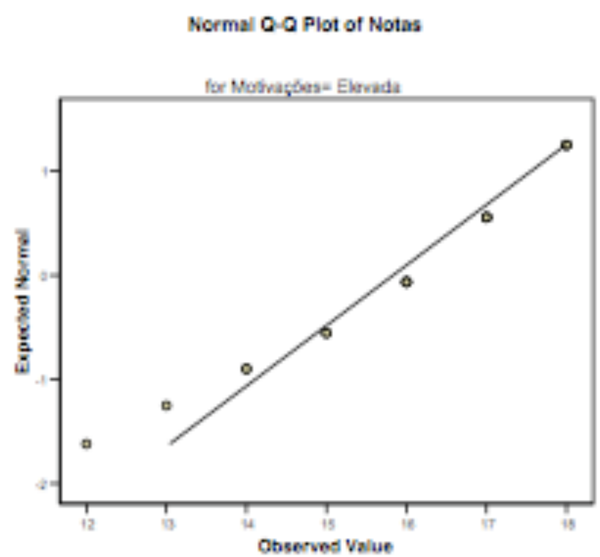
Verifica-se a normalidade para todos os grupos.

Exercício exemplo: Anova

Normalidade







Exercício exemplo: Anova

Análise de variâncias

$$\begin{cases} H_0: \sigma^2_{MBAIXA} = \sigma^2_{MMEDIA} = \sigma^2_{MALTA} \\ H_1: \sigma^2_i \neq \sigma^2_j \quad \text{para pelo menos um par } (i,j) \end{cases}$$

Test of Homogeneity of Variances

Notas

Levene Statistic	df1	df2	Sig.
1,073	2	51	,349

Como o p-value=0.349 > 0.05=α, NÃO se REJEITA H_0. (Não se rejeita a igualdade de variâncias).

Verificam-se as condições de aplicabilidade da ANOVA!

Exercício exemplo: Anova

Comparação de médias

Hipóteses:

$$H_0: \mu_{MBAIXA} = \mu_{MMEDIA} = \mu_{MALTA} = \mu$$

vs

$$H_1: \; {}_{(i,j)} : \; {}_{i} \qquad {}_{j} , \; i \quad j$$

i, j = baixa, média, alta

ANOVA

Notas

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	113,593	2	56,796	20,847	,000
Within Groups	138,944	51	2,724		
Total	252,537	53			

Exercício exemplo: Anova

Comparação de médias

ANOVA

Notas

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	113,593	2	56,796	20,847	,000
Within Groups	138,944	51	2,724		
Total	252,537	53			

2) $p = 0{,}000 < 0{,}05$ logo rejeita-se H_0, concluindo-se que existe pelo menos um grupo de alunos que difere de pelo menos outro, no que diz respeito às suas motivações.

Quando tal sucede, há que verificar ONDE SE SITUAM AS DIFERENÇAS.

COMPARAÇÕES MÚLTIPLAS

Exercício exemplo: Anova

Comparações Multiplas

Dependent Variable: Notas

	(I) Motivação	(J) Motivação	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Tukey HSD	1 Baixa	2 Média	-1,222	,550	,077	-2,55	,11
		3 Elevada	-3,500*	,550	,000	-4,83	-2,17
	2 Média	1 Baixa	1,222	,550	,077	-,11	2,55
		3 Elevada	-2,278*	,550	,000	-3,61	-,95
	3 Elevada	1 Baixa	3,500*	,550	,000	2,17	4,83
		2 Média	2,278*	,550	,000	,95	3,61
Scheffe	1 Baixa	2 Média	-1,222	,550	,095	-2,61	,17
		3 Elevada	-3,500*	,550	,000	-4,89	-2,11
	2 Média	1 Baixa	1,222	,550	,095	-,17	2,61
		3 Elevada	-2,278*	,550	,001	-3,67	-,89
	3 Elevada	1 Baixa	3,500*	,550	,000	2,11	4,89
		2 Média	2,278*	,550	,001	,89	3,67
LSD	1 Baixa	2 Média	-1,222*	,550	,031	-2,33	-,12
		3 Elevada	-3,500*	,550	,000	-4,60	-2,40
	2 Média	1 Baixa	1,222*	,550	,031	,12	2,33
		3 Elevada	-2,278*	,550	,000	-3,38	-1,17
	3 Elevada	1 Baixa	3,500*	,550	,000	2,40	4,60
		2 Média	2,278*	,550	,000	1,17	3,38
Bonferroni	1 Baixa	2 Média	-1,222	,550	,092	-2,58	,14
		3 Elevada	-3,500*	,550	,000	-4,86	-2,14
	2 Média	1 Baixa	1,222	,550	,092	-,14	2,58
		3 Elevada	-2,278*	,550	,000	-3,64	-,92
	3 Elevada	1 Baixa	3,500*	,550	,000	2,14	4,86
		2 Média	2,278*	,550	,000	,92	3,64

*. The mean difference is significant at the .05 level.

À excepção da motivação (Baixa/Média), todos os testes acusam diferenças entre motivações.

EXERCICIOS 1 E 2

EXEL & IBM SPSS