Teste

de

Hipóteses

jlborges@fe.up.pt

jlborges@fe.up.pt

O objectivo de um **teste de hipóteses**

é o de verificar se uma estimativa pontual obtida a partir de uma amostra é ou não compatível com uma determinada população

O resultado de um teste é do tipo **afirmativo / negativo**, havendo a possibilidade de **erro**

Exemplo:
Como podemos verificar se uma moeda está, ou não, viciada?

Como podemos verificar se uma moeda está, ou não, viciada?

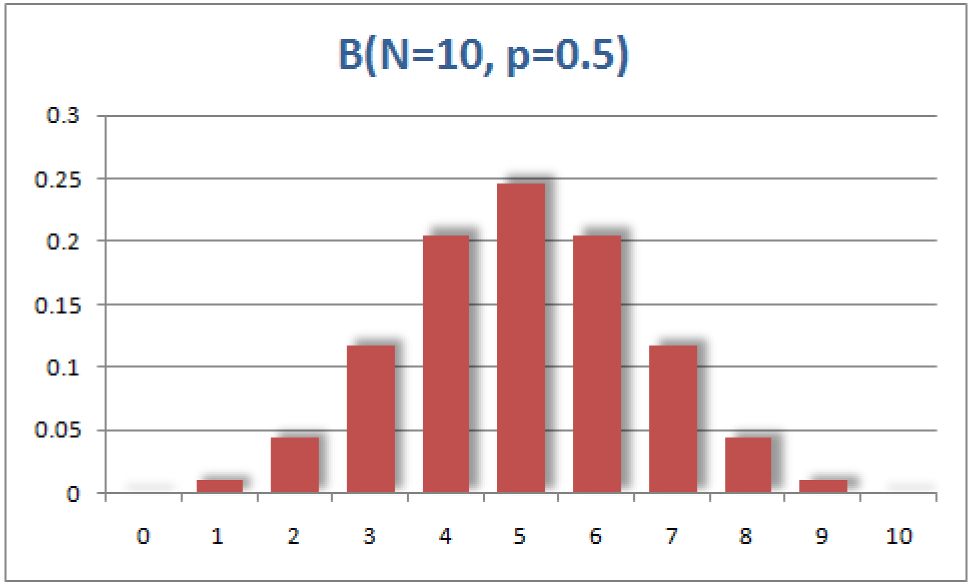
Vamos lançar várias vezes a moeda e verificar se o resultado obtido é plausível para moedas não viciadas.

Vamos então lançar a moeda 10 vezes.

Se obtiver 4 caras qual seria a conclusão?

e se obtiver 8 caras?

y	p(y)
0	0.001
1	0.010
2	0.044
3	0.117
4	0.205
5	0.246
6	0.205
7	0.117
8	0.044
9	0.010
10	0.001



Procedimento Básico de Teste de Hipóteses

1. Definição das hipóteses
2. Identificação da estatística de teste e caracterização da sua distribuição
3. Estabelecimento da regra de decisão, com especificação do nível de significância do teste
4. Cálculo da estatística de teste e tomada de decisão
5. Cáculo do valor de prova

De acordo com o eurobarometro a altura média dos Portugueses é 165 cm com um desvio padrão de 7 cm

Parece-me que os alunos da FEUP são mais altos do que isso

Há cerca de 7000 alunos na FEUP

Como posso verificar se isso é verdade?

1. Definição das Hipóteses

Hipótese alternativa ou H_1

Hipótese que traduz uma conjectura que se pretende verificar

Contém **sempre uma desigualdade** (> ou <) ou uma **não-igualdade** (≠) e nunca uma igualdade (=)

	Teste bilateral	Teste unilateral à esquerda	Teste unilateral à direita
H_1	$\neq$	$<$	$>$

Definição das Hipóteses

- **Hipótese nula ou H_0**

H_0 é considerada verdadeira até prova em contrário
(se H_0 for provada como falsa é rejeitada e podemos aceitar H_1 como válida)

Hipótese complementar de H_1

Contém **sempre uma igualdade** (=, ≥, ≤),
apenas se testando a situação de "=" por ser a que mais se aproxima de H_1

Definição das Hipóteses

Hipótese Nula:	Traduz a hipótese que vai ser testada
Hipótese Alternativa:	Hipótese alternativa à hipótese nula

Exemplo:

O responsável pelas tecnologias de informação de uma empresa pretende avaliar a capacidade de transmissão de dados da Intranet.

Um dos testes a efectuar consiste em estimar o tempo médio de transferência de um ficheiro de 250 Mb, sendo a finalidade do teste **validar a conjectura de que o tempo médio é inferior a 26 segundos**.

$H_0 : \mu \geq 26$ ou $\mu = 26$ Complementar a H1. Igualdade.

$H_1 : \mu < 26$ Aquilo que se pretende verificar. Desigualdade.

- ***Intuição:***
  - **Consideramos que μ=26 é verdadeira.**
  - **Vamos recolher uma amostra e calcular o tempo médio.**
  - **Se o resultado obtido for muito improvável para μ=26 rejeitamos essa hipótese e consideramos como válida a hipótese complementar.**

  - Pergunta: O que é um resultado pouco provável?

jlborges@fe.up.pt

Exemplo

Para **verificar a conjectura acerca do tempo médio de transferência** ($\mu < 26$) foi **recolhida uma amostra aleatória do tempo de 50 transferências**. Admita que se conhece o desvio padrão populacional σ=1.84,

27.3	26.8	23.8	23.5	25.3	26.0	25.4	26.4	25.1	25.4
27.0	29.4	25.5	24.1	25.1	25.3	23.8	23.4	26.2	26.2
25.9	25.5	24.2	23.7	27.6	23.6	25.1	24.8	27.8	25.1
24.0	26.9	21.2	24.3	24.6	26.0	24.1	22.6	23.3	25.3
29.7	28.3	27.0	22.2	26.0	25.4	25.8	26.0	23.8	21.5

$\bar{x} = 25.51 \quad s^2 = 1.65^2$

Será este resultado pouco provável?

2. Identificação da Estatística de Teste e Caracterização da Sua Distribuição

A **estatística de teste (ET)** é utilizada para verificar a plausibilidade de H_0 sendo necessário **conhecer a sua distribuição** quando se admite que H_0 é verdadeira.

Exemplo (cont.)

A estatística que corresponde à média populacional, μ, é a média amostral. Uma vez que N=50 qualquer que seja a distribuição de *x* temos que

$$\bar{x} \rightarrow N\left(\mu, \sigma^2 / N\right)$$

e, como o desvio padrão populacional é conhecido, temos que

$$\boxed{ET = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{N}} \rightarrow N(0,1)}$$

$$\underbrace{\boxed{ET = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{N}} \quad \to \quad N(0,1)}}$$

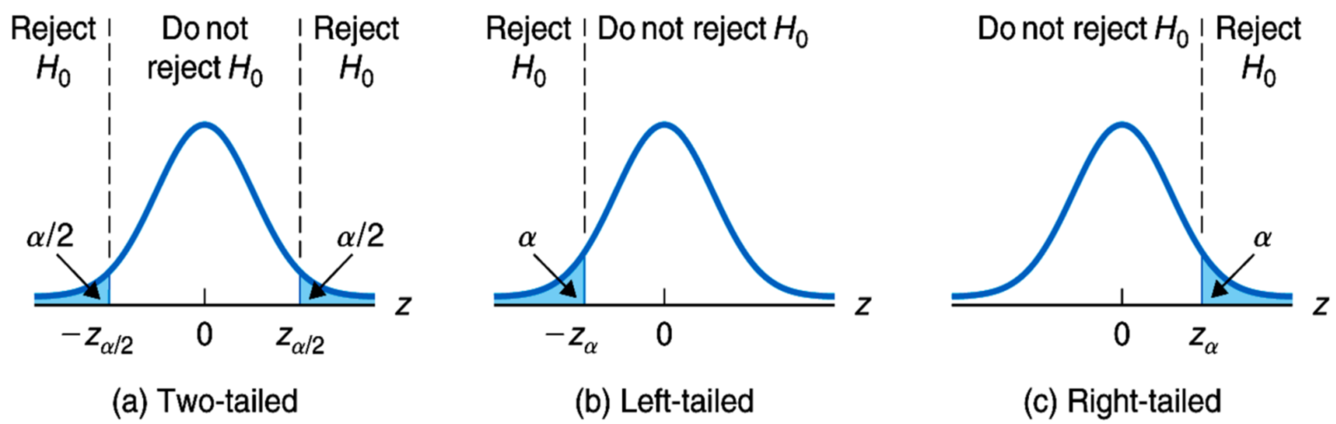
Como sabemos qual a sua distribuição, a ET permite-nos verificar se se trata de um resultado pouco provável, ou não.

3. Estabelecimento da Regra de Decisão e Especificação do Nível de Significância do Teste

A hipótese nula será rejeitada se o valor de **ET** for muito improvável face à sua distribuição quando H_0 é verdadeira.

A regra de decisão fixa o valor **ET(α)**, valor crítico, a partir do qual se rejeita H_0, criando uma região de rejeição.

A probabilidade α designa-se nível de significância do teste.



A probabilidade α designa-se nível de significância do teste.

- ***Intuição***
  - **Numa distribuição Normal todos os resultados são possíveis, uma vez que a distribuição não é limitada superiormente ou inferiormente.**
  - **Logo, temos de estabelecer um limite a partir do qual consideramos os resultados demasiado improváveis.**
  - **Esse limite é o α e corresponde à probabilidade de tomar uma decisão errada.**

- No exemplo do lançamento de uma moeda 10 vezes consideramos muito pouco provável obter 9 ou 10 caras. No entanto esse resultado é, de facto, possível.

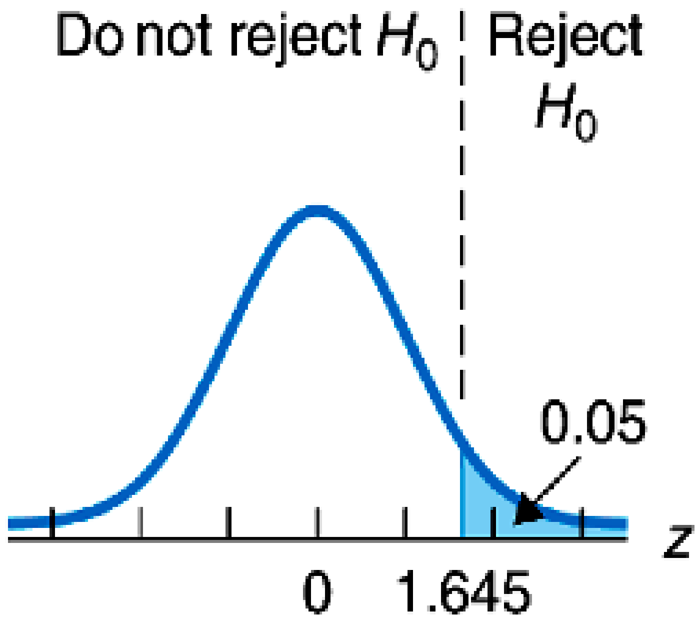
jlborges@fe.up.pt

O nível de significância do teste, α, representa a probabilidade de se rejeitar H_0 quando esta hipótese é verdadeira.

Este erro designa-se por **Erro do Tipo I**.

Valores mais comuns de Z_α

$z_{0.10}$	$z_{0.05}$	$z_{0.025}$	$z_{0.01}$	$z_{0.005}$
1.28	1.645	1.96	2.33	2.575

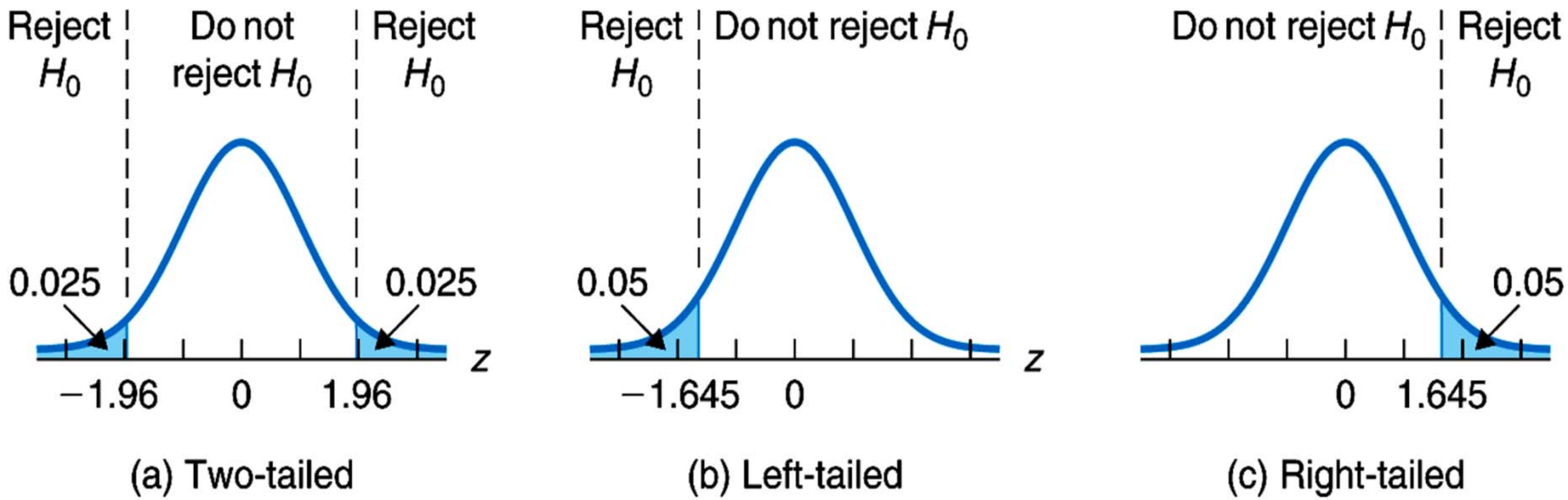


(c) Right-tailed

jlborges@fe.up.pt

Quando

$$ET \to N(0,1) \quad e \quad \alpha = 5\%$$



Exemplo (cont.)

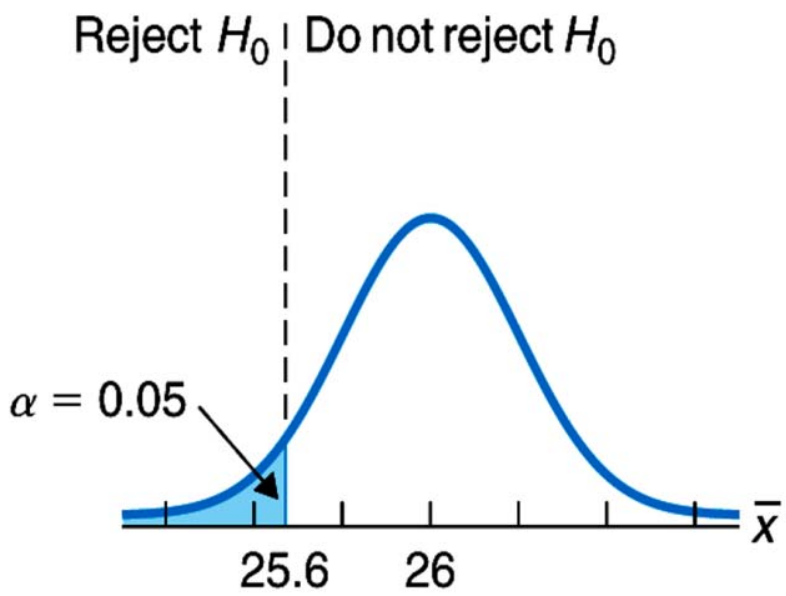
$$N = 50 \quad \bar{x} = 25.51 \quad \sigma^2 = 1.84^2$$

$$H_0 : \mu \geq 26 \text{ ou } \mu = 26$$

$$z(\alpha) = -1.645$$

$$\frac{x_{crt} - \mu}{\sigma / \sqrt{N}} = -1.645$$

$$x_{crt} = -1.645 \cdot \frac{1.84}{\sqrt{50}} + 26 = 25.57$$



Se o valor amostral for inferior ao Valor Crítico a hipótese nula deverá ser rejeitada.

Exemplo (cont.)

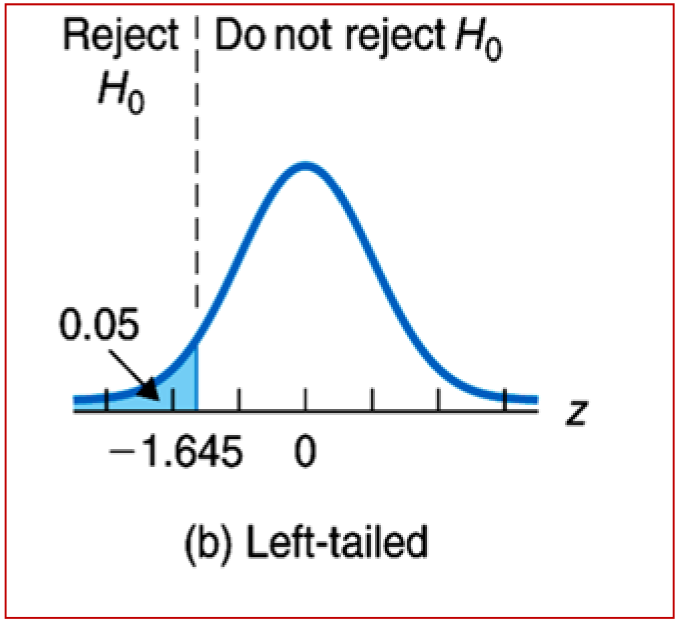
Regra de decisão:

Rejeitar se ET < -1.645

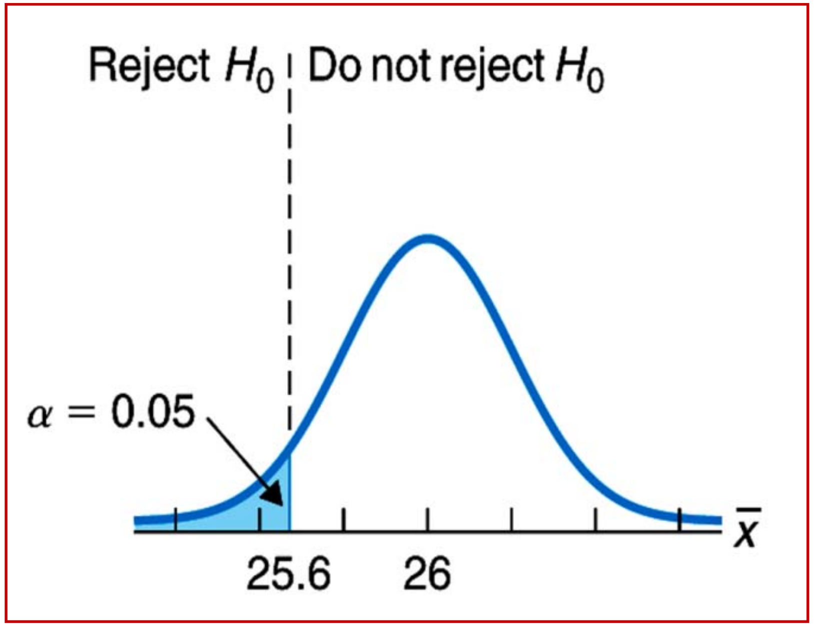
ou então

Rejeitar se X < 25.6

***As duas formas são equivalentes*.**



(b) Left-tailed



4. Cálculo da Estatística de Teste e Tomada de Decisão

Corresponde ao cálculo do valor de **ET** e consequente aplicação da regra de decisão

Exemplo (cont.)

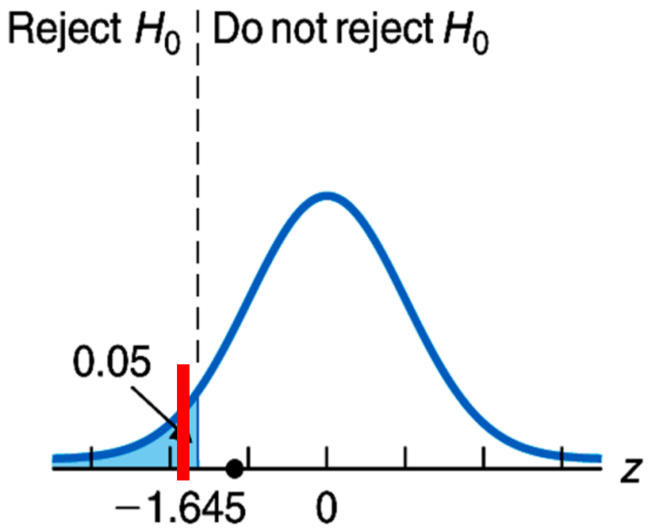
$$\alpha = 5\% \quad \Rightarrow \quad ET(\alpha) = -1.645$$

Regra de decisão

$ET < -1.645$ rejeitar H_0 e aceitar H_1

$ET \geq -1.645$ não rejeitar H_0 (teste inconclusivo)

$$ET = \frac{\overline{x} - \mu}{\sigma / \sqrt{N}} = \frac{25.51 - 26}{1.84 / \sqrt{50}} = -1.88$$



REJEITAR H0

e aceitar H1

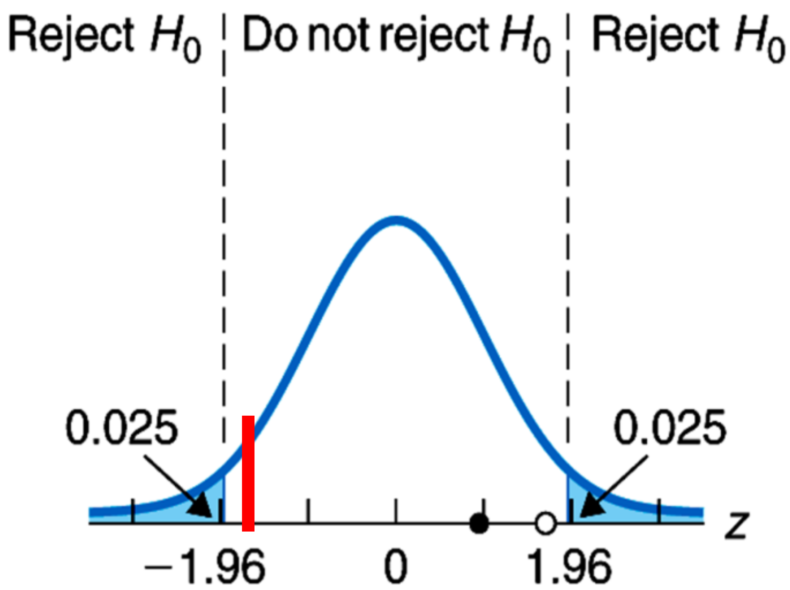
Exemplo (cont.)

E se o teste fosse bilateral?

$$\alpha = 2.5\% \quad \Rightarrow \quad ET(\alpha) = -1.96$$

$$ET = \frac{\bar{x} - \mu}{\sigma / \sqrt{N}} = \frac{25.51 - 26}{1.84 / \sqrt{50}} = -1.88$$

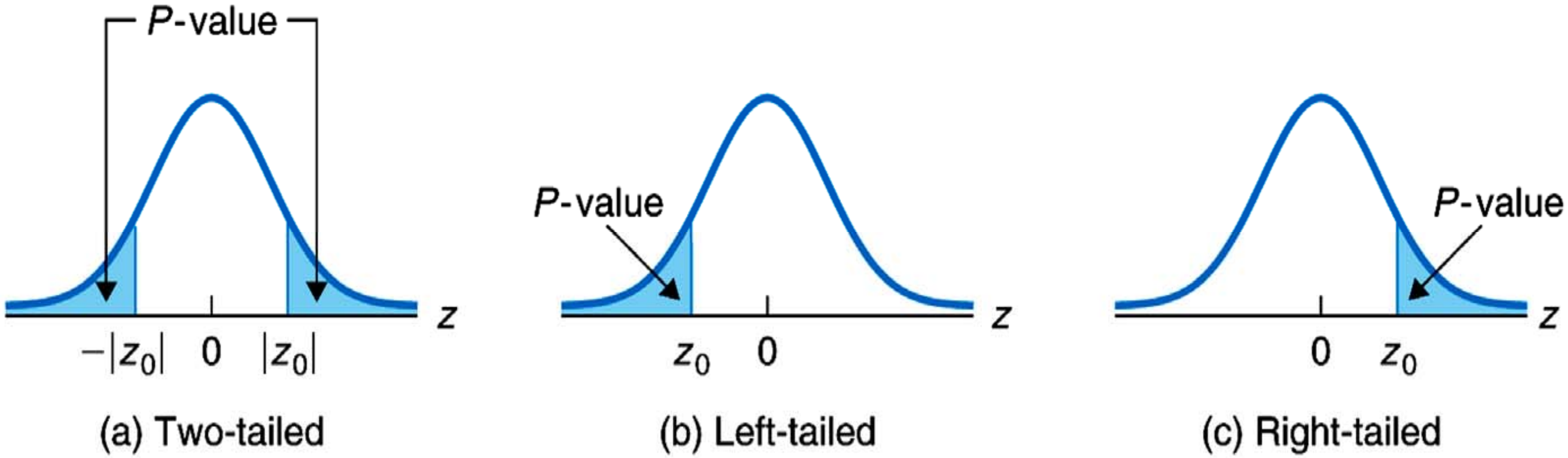
Não há evidência contra H0 $\rightarrow$ teste inconclusivo



Note que, para que o teste seja honesto, as hipóteses tem de ser definidas antes de calculada a estatística de teste.

Valor de Prova

O valor de prova, **valor P**, corresponde à probabilidade de **ET** tomar um valor igual ou mais extremo do que o registado.
Constitui uma medida do grau com que os dados amostrais contradizem H_0.



$$Z_0 \equiv ET$$

Permite tomar uma decisão mais fundamentada, porque em vez de ser do tipo rejeita - não rejeita (decisão binária) dá-nos uma medida da distância relativamente à decisão oposta.

Exemplo (cont.)

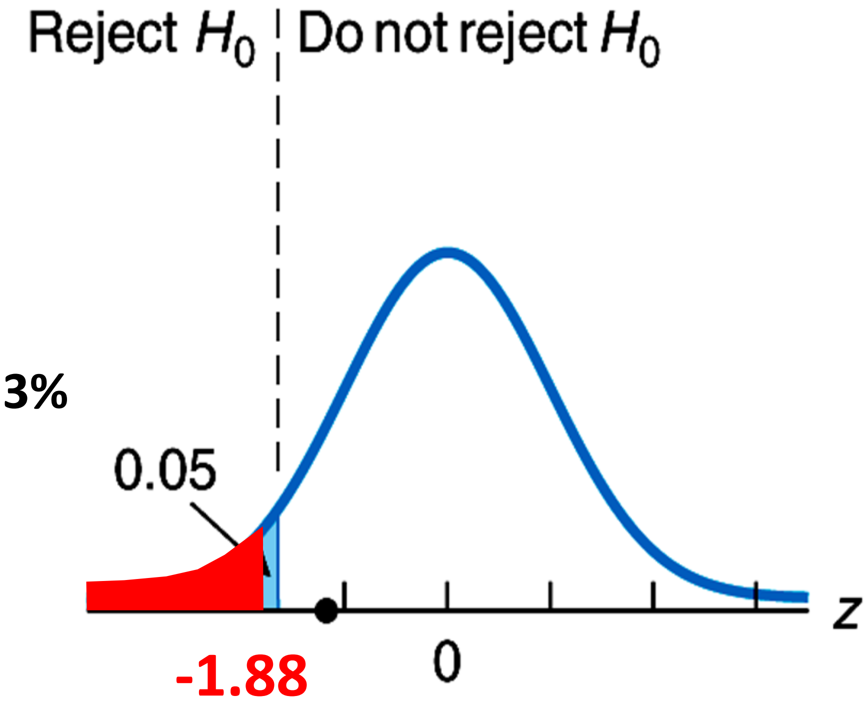
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.067	0.066	0.064	0.063	0.062	0.061	0.059	0.058	0.057	0.056
1.6	0.055	0.054	0.053	0.052	0.051	0.050	0.049	0.048	0.047	0.046
1.7	0.045	0.044	0.043	0.042	0.041	0.040	0.039	0.038	0.038	0.037
1.8	0.036	0.035	0.034	0.034	0.033	0.032	0.031	0.031	0.030	0.029
1.9	0.029	0.028	0.027	0.027	0.026	0.026	0.025	0.024	0.024	0.023
2.0	0.023	0.022	0.022	0.021	0.021	0.020	0.020	0.019	0.019	0.018

$$ET = \frac{\bar{x} - \mu}{\sigma / \sqrt{N}} =$$

$$\frac{25.51 - 26}{1.84 / \sqrt{50}} = -1.88$$

$$\text{valorP} = P\left(ET \leq -1.88 \mid H_0\right) = 0.03$$

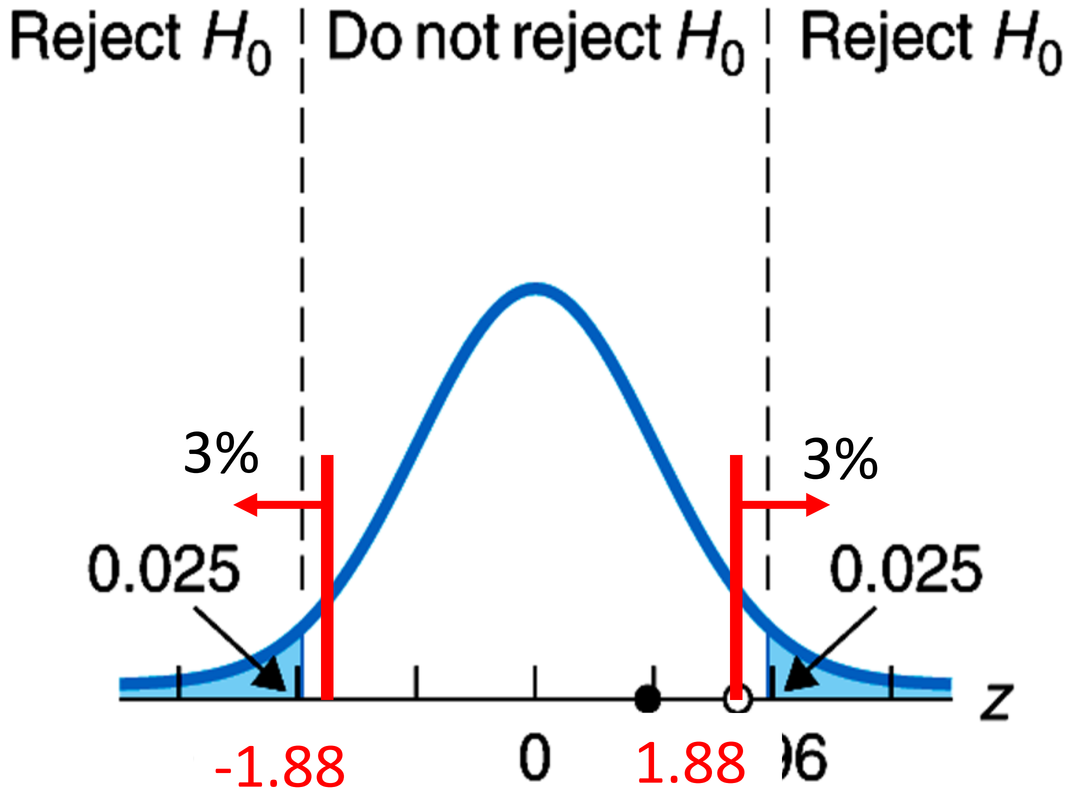
Valor de Prova = 3%



Exemplo (cont.)

Caso do teste tivesse sido definido como <u>bilateral</u> (H_1: $\mu \neq 26$)

$$\text{valor P} = P\left[\left(ET \leq -1.88 \mid H_0\right) \cup \left(ET \geq 1.88 \mid H_0\right)\right] = 2 \cdot 0.03 = 0.06$$



O valor de prova dá uma medida da evidência contra H0.

A decisão deve ser tomada em função do risco que estamos dispostos a cometer.

É diferente estar a avaliar o tempo médio de transferência de um ficheiro ou a tensão de rotura de uma asa de um avião.

P-value	Evidence against H_0
$P > 0.10$	Weak or none
$0.05 < P \leq 0.10$	Moderate
$0.01 < P \leq 0.05$	Strong
$P \leq 0.01$	Very strong

jlborges@fe.up.pt

Testes ao Valor Esperado

amostra grande, população qualquer	$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0 \;\; \mu < \mu_0 \;\; \mu \neq \mu_0$	$ET = \dfrac{\overline{X} - \mu_0}{S / \sqrt{N}} \rightarrow N(0,1)$
amostra pequena, população normal	$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0 \;\; \mu < \mu_0 \;\; \mu \neq \mu_0$	$ET = \dfrac{\overline{X} - \mu_0}{S / \sqrt{N}} \rightarrow t_{N-1}$

Exemplo:

Com base na amostra (n=50) de idades de funcionários de uma determinada empresa **poder-se-á afirmar que a idade média dos funcionários da empresa é inferior a 39 anos**?

22	58	40	42	43	32	34	45	38	19
33	16	49	29	30	43	37	19	21	62
60	41	28	35	37	51	37	65	57	26
27	31	33	24	34	28	39	43	26	38
42	40	31	34	38	35	29	33	32	33

$$\bar{x} = 36.4$$

$$s = 12.1$$

$$H_0 : \mu \geq 39$$

Amostra de grande dimensão, teste Z.

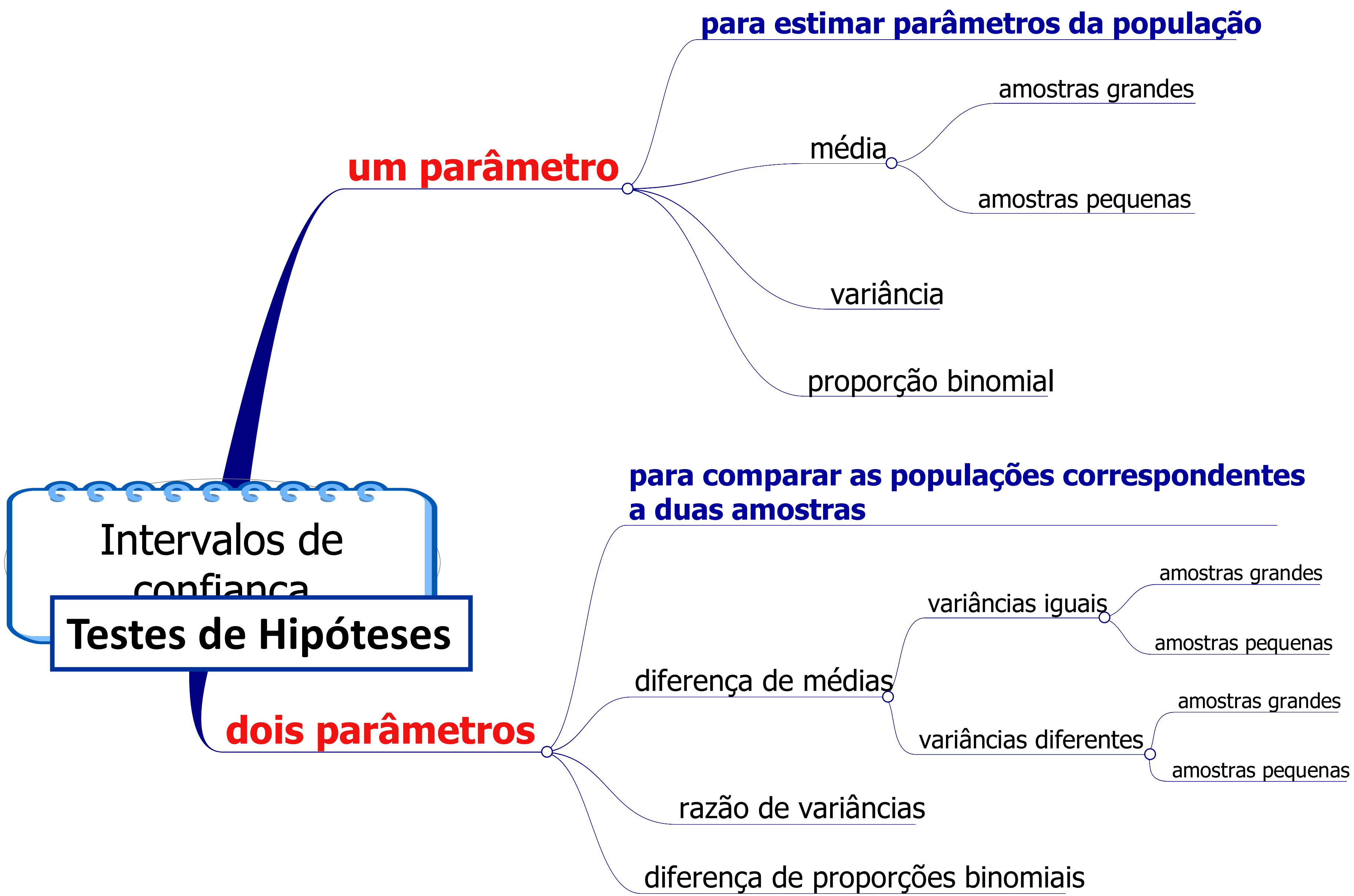
$$H_1 : \mu < 39$$

$$ET = \frac{36.4 - 39}{12.1/\sqrt{50}} = -1.52$$

$$\alpha = 5\% \quad \Rightarrow \quad z(\alpha) = -1.96$$

Não rejeitar H0

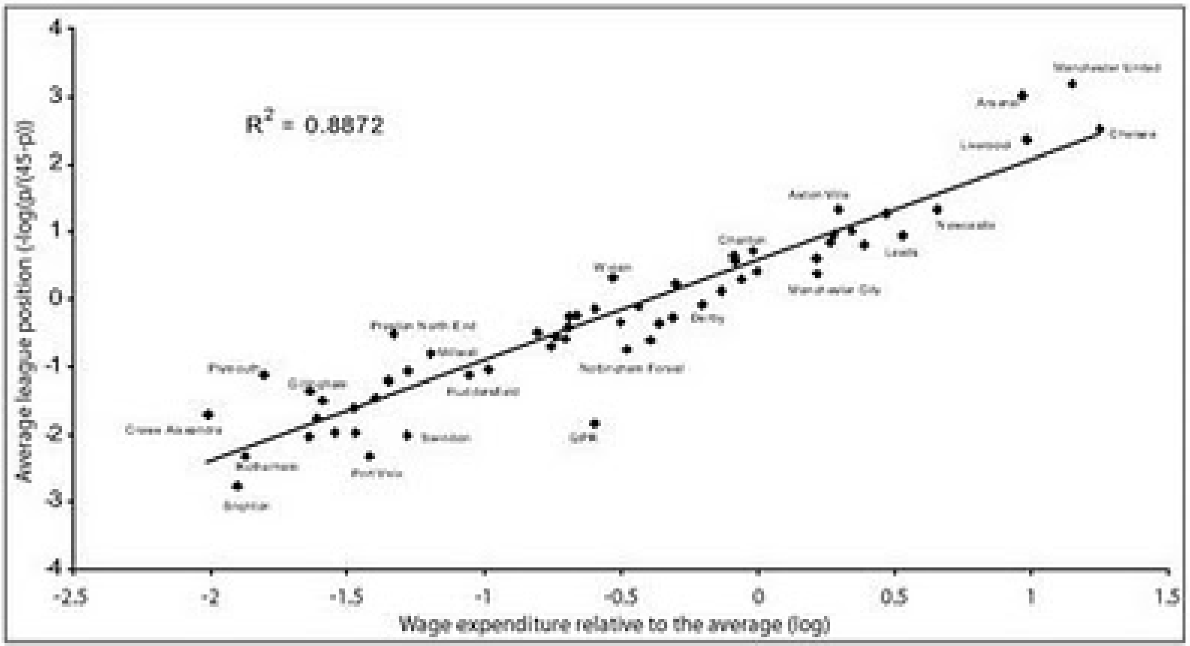
Valor de prova = p(Z<-1.52) = 6.4%

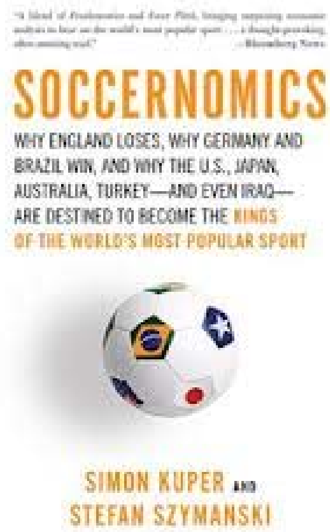

Intervalos de confiança
Testes de Hipóteses
um parâmetro
para estimar parâmetros da população
média
amostras grandes
amostras pequenas
variância
proporção binomial
dois parâmetros
para comparar as populações correspondentes a duas amostras
diferença de médias
variâncias iguais
amostras grandes
amostras pequenas
variâncias diferentes
amostras grandes
amostras pequenas
razão de variâncias
diferença de proporções binomiais

Variância de uma População Normal	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2\,,\quad \sigma^2 > \sigma_0^2 \quad ou \quad \sigma^2 < \sigma_0^2$	$ET = (N-1)\cdot\dfrac{S^2}{\sigma_0^2} \quad \rightarrow \quad \chi^2_{N-1}$
Razão de Variâncias de Duas Populações Normais	$H_0: \sigma_A^2 = \sigma_B^2 \equiv \dfrac{\sigma_A^2}{\sigma_B^2} = 1$	$ET = \dfrac{S_A^2}{S_B^2} \quad \rightarrow \quad F_{N_{A-1},N_{B-1}}$
Valor Esperado de uma População (amostra grande, população qualquer)	$H_0: \mu = \mu_0$	$ET = \dfrac{\overline{X} - \mu_0}{S/\sqrt{N}} \quad \rightarrow \quad N(0,1)$
amostra pequena, população normal		$ET = \dfrac{\overline{X} - \mu_0}{S/\sqrt{N}} \quad \rightarrow \quad t_{N-1}$
Diferença de Valores Esperados de duas Populações (amostras grandes, independentes, populações quaisquer)	$H_0: \mu_A - \mu_B = \delta_0$	$ET = \dfrac{(\overline{X}_A - \overline{X}_B) - \delta_0}{\sqrt{S_A^2/N_A + S_B^2/N_B}} \rightarrow N(0,1)$
	$S^2 = \dfrac{(N_A - 1)\cdot S_A^2 + (N_B - 1)\cdot S_B^2}{N_A + N_B - 2}$	$ET = \dfrac{(\overline{X}_A - \overline{X}_B) - \delta_0}{S\cdot\sqrt{1/N_A + 1/N_B}} \rightarrow N(0,1)$
amostras pequenas, independentes, populações normais	$ET = \dfrac{(\overline{X}_A - \overline{X}_B) - \delta_0}{\sqrt{S_A^2/N_A + S_B^2/N_B}} \rightarrow t_{GL}$	$ET = \dfrac{(\overline{X}_A - \overline{X}_B) - \delta_0}{S\cdot\sqrt{1/N_A + 1/N_B}} \rightarrow t_{N_A + N_B - 2}$
Proporção Binomial (amostra de grande dimensão)	$H_0: p = p_0$	$ET = \dfrac{Y/N - p_0}{\sqrt{(p_0\cdot(1 - p_0))/N}} \rightarrow N(0,1)$
Diferença de Duas Proporções Binomiais (amostras de grande dimensão)	$H_0: p_A - p_B = p_0$	$ET = \dfrac{(Y_A/N_A - Y_B/N_B) - p_0}{\sqrt{Y_A\cdot(N_A - Y_A)/N_A^3 + Y_B\cdot(N_B - Y_B)/N_B^3}} \quad \rightarrow \quad N(0,1)$

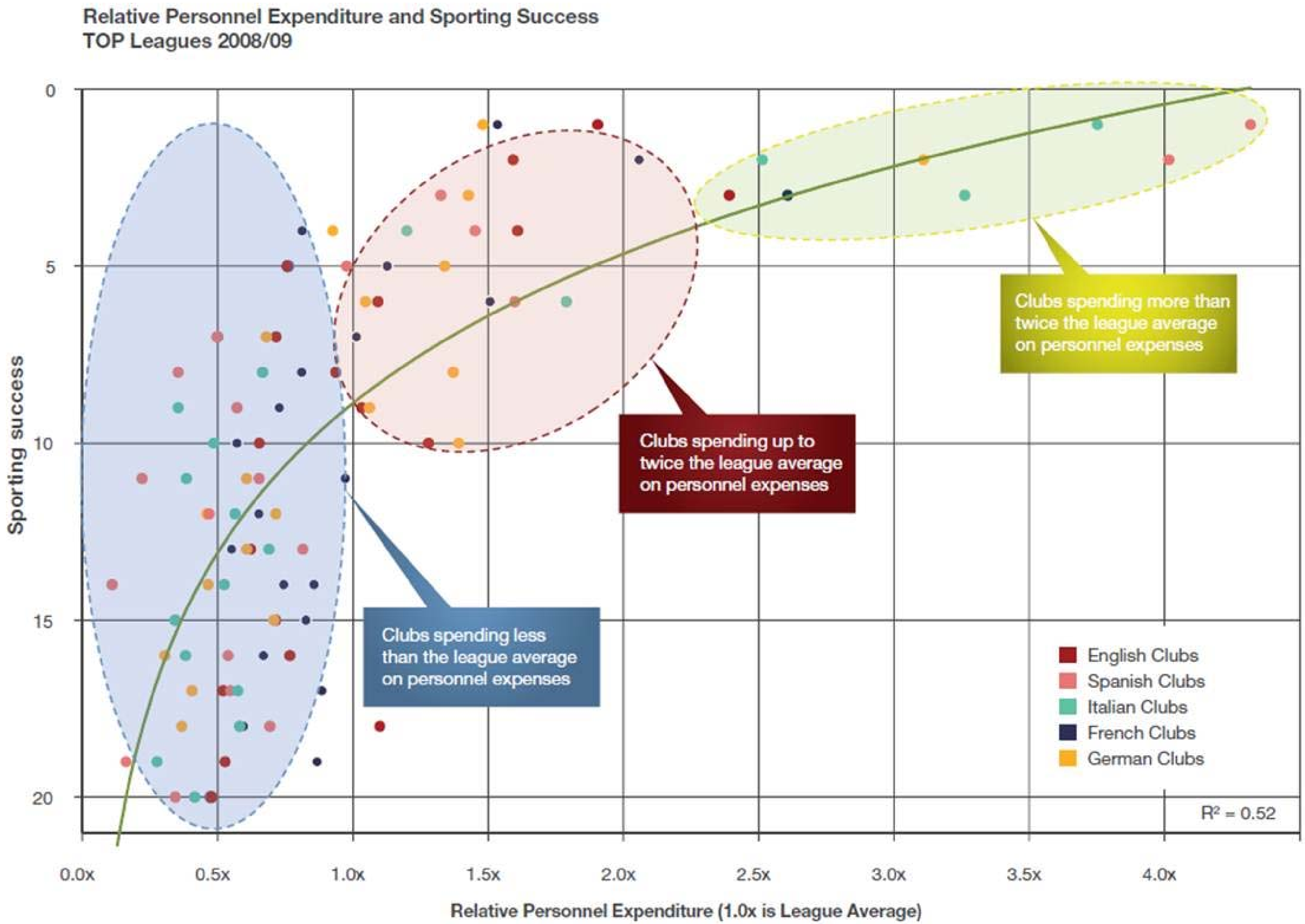
Quantifying the Value of a Football Manager

- In the Soccernomics book, the authors Simon Kuper and Stefan Szymanski lay out a very intuitive yet startling correlation:
  - To **finish higher in the tables** of the top two English soccer leagues, one must **spend more money than their opponents**.
  - Their analysis of the datashows that the team payroll as a function of a multiple of the leagues' average payroll explains a 88.7% of the variation in final classification.





http://leastthing.blogspot.pt/2012/01/quantifying-value-of-football-manager.html



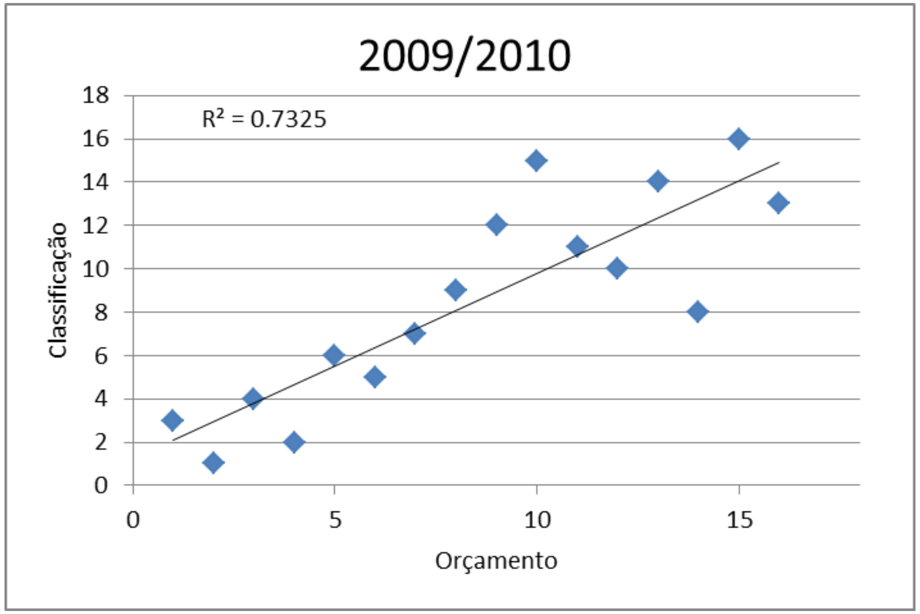
2009/2010						2010/2011						2011/2012				
Equipa	**Orc.**	**Rk_O**	**Pts**	**Rk_P**		**Equipa**	**Orc.**	**Rk_O**	**Pts**	**Rk_P**		**Equipa**	**Orc.**	**Rk_O**	**Pts**	**Rk_P**
FC Porto	40	1	68	3		FC Porto	94.7	1	84	1		FC Porto	95	1	75	1
Benfica	30	2	76	1		Benfica	28.43	2	63	2		Benfica	50	2	69	2
Sporting	22.5	3	48	4		Guimarães	11	5	43	5		Sporting	40	3	59	4
Braga	9	4	71	2		Sporting	14.5	4	48	3		Braga	17	4	62	3
Marítimo	7	5	41	6		Braga	17	3	46	4		Guimarães	9	5	45	6
Guimarães	6.9	6	41	5		Marítimo	3.5	9	35	10		Maritimo	6	6	50	5
Nacional	5	7	39	7		Beira Mar	1.8	14	33	13		Nacional	5	7	44	7
Leiria	4	8	35	9		Portimonense	1.2	16	25	15		Leiria	4.5	8	19	16
Rio Ave	4	9	31	12		Nacional	4	8	42	6		Rio Ave	4	9	28	14
Belenenses	3.7	10	23	15		Rio Ave	3.45	10	38	8		Gil Vicente	3.7	10	34	9
Académica	3.6	11	33	11		Leiria	4	7	35	9		Académica	3.5	11	29	13
Setúbal	2.5	13	25	14		Setúbal	2	13	34	11		Paços de Ferreira	2.5	12	31	10
Paços de Ferreira	2.5	12	35	10		Olhanense	1.4	15	34	12		Beira-Mar	2	13	29	12
Naval	2.4	14	36	8		Paços de Ferreira	2.5	12	41	7		Feirense	2	14	24	15
Leixões	2	15	21	16		Naval	2.8	11	23	16		Olhanense	1.4	15	39	8
Olhanense	1.2	16	29	13		Académica	5.2	6	30	14		Setúbal	1.3	16	30	11

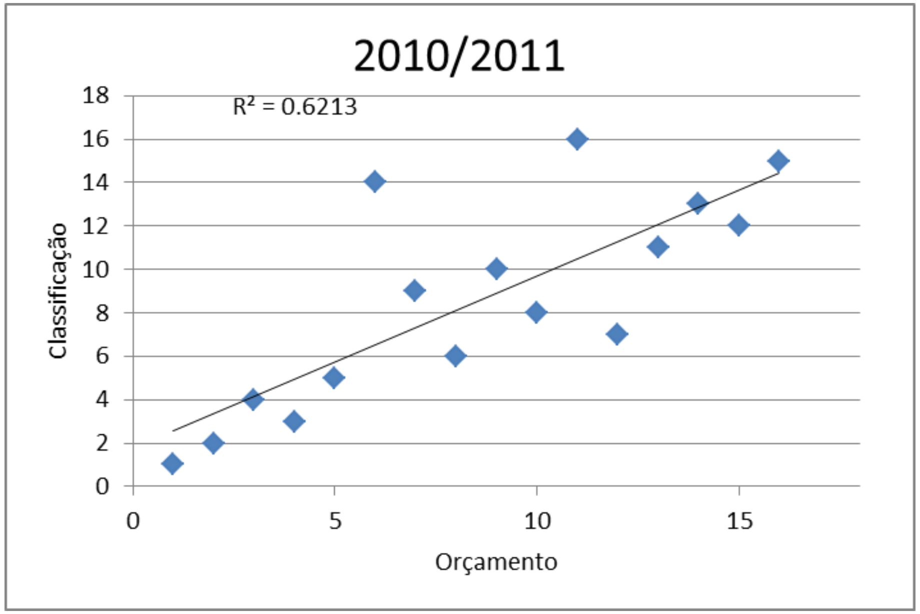
Fontes:

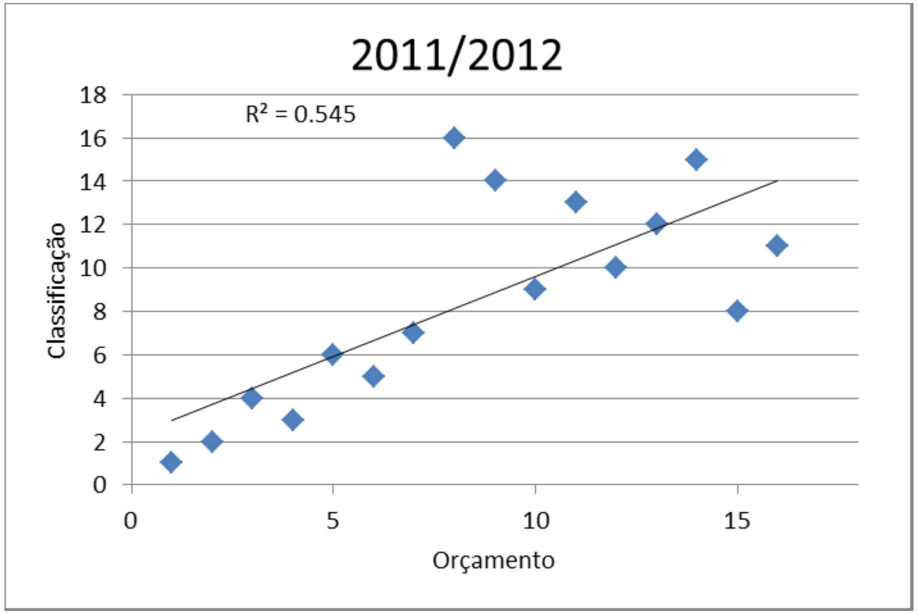
http://eusouogatomaltes.blogspot.pt/2010/05/1-liga-orcamentos-vs-classificacao.html

http://www.futebolive.tv/2011/05/orcamentos-ditam-resultados.html

http://desportoaveiro.blogs.sapo.pt/3398724.html


2009/2010
$R^2 = 0.7325$
Classificação
Orçamento
0
2
4
6
8
10
12
14
16
18
0
5
10
15

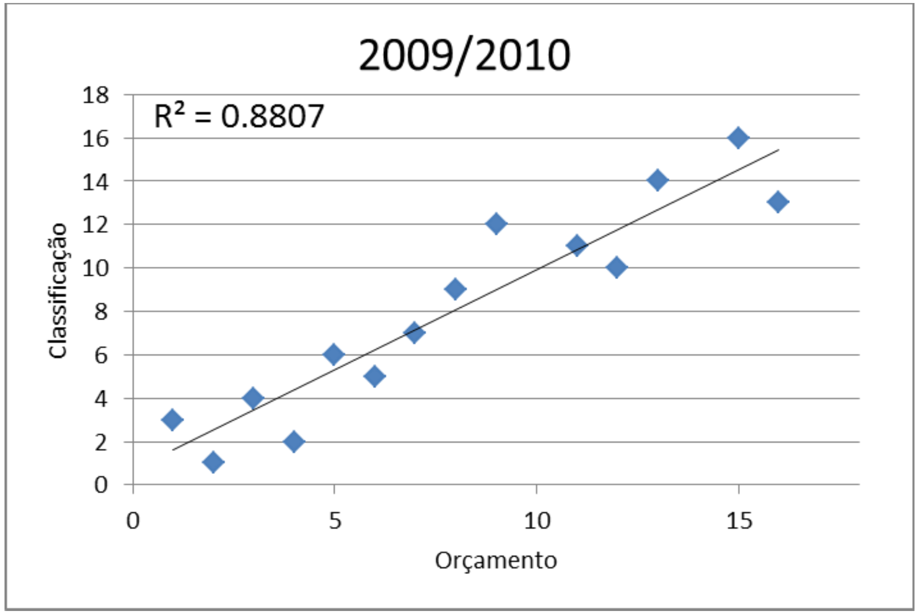

2010/2011
$R^2 = 0.6213$
Classificação
Orçamento
0
2
4
6
8
10
12
14
16
18
0
5
10
15

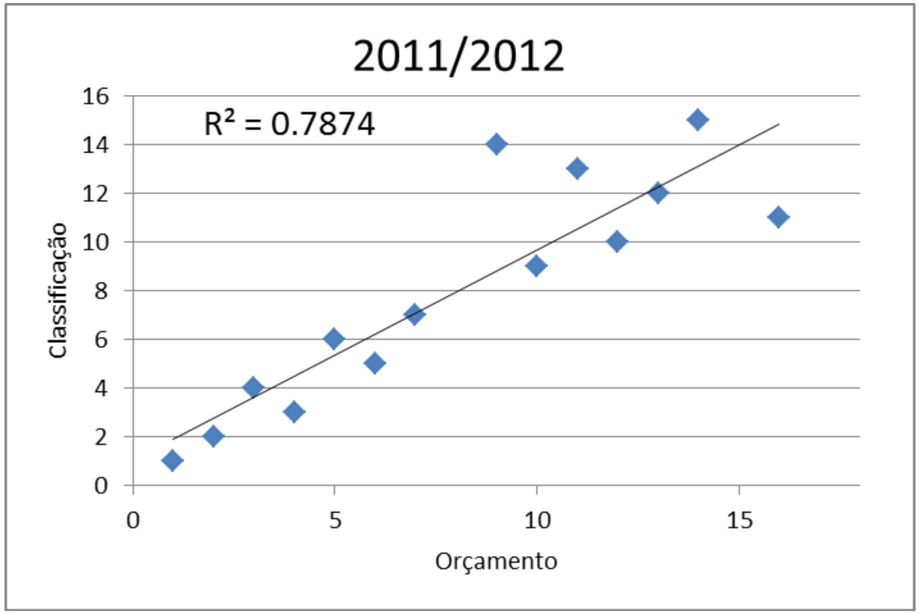

2011/2012
$R^2 = 0.545$
Classificação
Orçamento
0
2
4
6
8
10
12
14
16
18
0
5
10
15

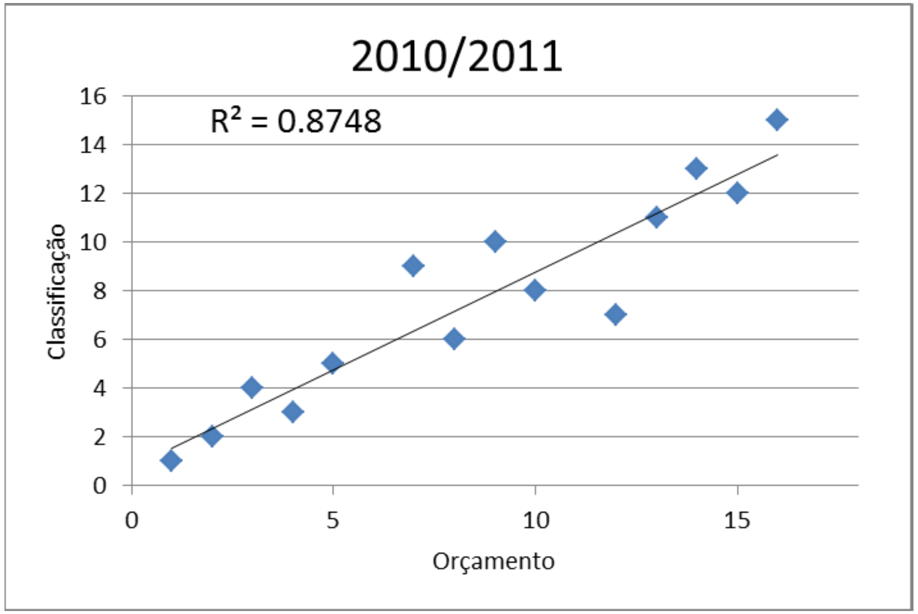
2009/2010					
Equipa	**Orc.**	**Rk_O**	**Pts**	**Rk_P**	**diff**
Nacional	5	7	39	7	0
Académica	3.6	11	33	11	0
Benfica	30	2	76	1	1
Sporting	22.5	3	48	4	1
Marítimo	7	5	41	6	1
Guimarães	6.9	6	41	5	1
Leiria	4	8	35	9	1
Setúbal	2.5	13	25	14	1
Leixões	2	15	21	16	1
FC Porto	40	1	68	3	2
Braga	9	4	71	2	2
Paços de Fe	2.5	12	35	10	2
Rio Ave	4	9	31	12	3
Olhanense	1.2	16	29	13	3
Belenenses	3.7	10	23	15	5
Naval	2.4	14	36	8	6

2010/2011					
Equipa	**Orc.**	**Rk_O**	**Pts**	**Rk_P**	**diff**
FC Porto	94.7	1	84	1	0
Benfica	28.43	2	63	2	0
Guimarães	11	5	43	5	0
Sporting	14.5	4	48	3	1
Braga	17	3	46	4	1
Marítimo	3.5	9	35	10	1
Beira Mar	1.8	14	33	13	1
Portimonen	1.2	16	25	15	1
Nacional	4	8	42	6	2
Rio Ave	3.45	10	38	8	2
Leiria	4	7	35	9	2
Setúbal	2	13	34	11	2
Olhanense	1.4	15	34	12	3
Paços de Fe	2.5	12	41	7	5
Naval	2.8	11	23	16	5
Académica	5.2	6	30	14	8

2011/2012					
Equipa	**Orc.**	**Rk_O**	**Pts**	**Rk_P**	**diff**
FC Porto	95	1	75	1	0
Benfica	50	2	69	2	0
Nacional	5	7	44	7	0
Braga	17	4	62	3	1
Sporting	40	3	59	4	1
Maritimo	6	6	50	5	1
Guimarães	9	5	45	6	1
Gil Vicente	3.7	10	34	9	1
Beira-Mar	2	13	29	12	1
Feirense	2	14	24	15	1
Paços de Fe	2.5	12	31	10	2
Académica	3.5	11	29	13	2
Setúbal	1.3	16	30	11	5
Rio Ave	4	9	28	14	5
Olhanense	1.4	15	39	8	7
Leiria	4.5	8	19	16	8


2009/2010
R² = 0.8807
Classificação
Orçamento


2011/2012
R² = 0.7874
Classificação
Orçamento


2010/2011
R² = 0.8748
Classificação
Orçamento

Na aula passada

Como podemos verificar se uma moeda está, ou não, viciada?

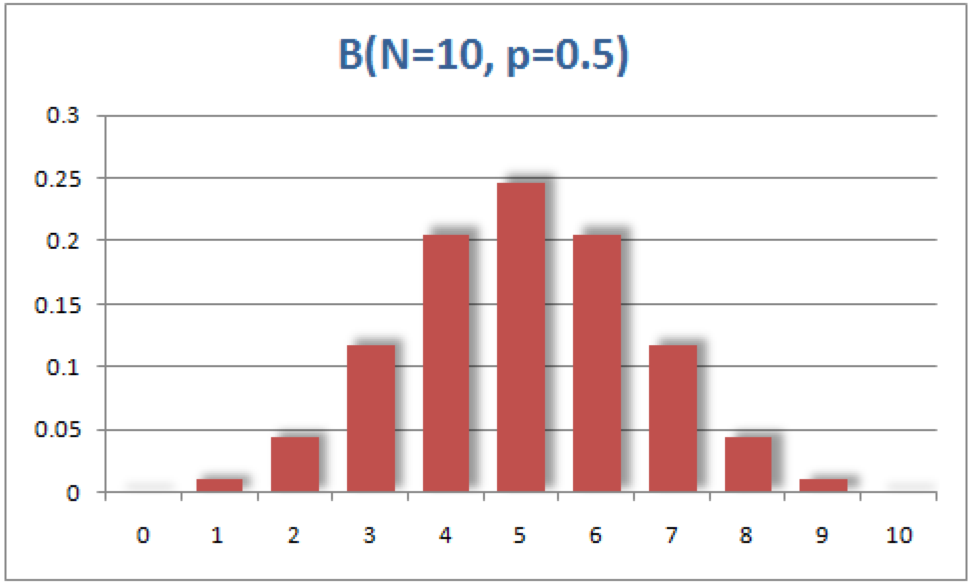
Vamos lançar várias vezes a moeda e verificar se o resultado obtido é plausível para moedas não viciadas.

Vamos então lançar a moeda 10 vezes.

Se obtiver 4 caras qual seria a conclusão?

e se obtiver 8 caras?

y	p(y)
0	0.001
1	0.010
2	0.044
3	0.117
4	0.205
5	0.246
6	0.205
7	0.117
8	0.044
9	0.010
10	0.001



Procedimento Básico de Teste de Hipóteses

1. Definição das hipóteses
2. Identificação da estatística de teste e caracterização da sua distribuição
3. Estabelecimento da regra de decisão, com especificação do nível de significância do teste
4. Cálculo da estatística de teste e tomada de decisão
5. Cáculo do valor de prova

jlborges@fe.up.pt

EXEMPLO

Uma amostra aleatória de 80 associados de um clube automóvel revelou que o montante gasto em prémios de seguro automóvel tem um valor médio de 800 Euros.

Verifique **se é razoável afirmar que o valor médio** gasto pelos sócios do clube em prémios de seguros **é superior a 770 Euros**.

$n = 80 \quad \bar{x} = 800$

$s = 200$

Definição das hipóteses

$$H_0 : \mu \leq 770$$

$$H_1 : \mu > 770$$

Aquilo que pretendemos verificar

$N = 80 \quad \bar{x} = 800 \quad \sigma^2 = 200^2$

$H_0 : \mu \leq 770$

$H_1 : \mu > 770$

Amostra grande $\Rightarrow$ ET $\rightarrow$ N(0,1)

$$ET = \frac{\bar{x} - \mu}{s / \sqrt{N}} = \frac{800 - 770}{200 / \sqrt{80}} = 1.34$$

$z(0.05) = 1.645$

$1.34 < 1.645 \quad \Rightarrow \quad$ não rejeitar H_0

Valor de Prova $= P\left(ET \geq 1.34 \mid H_0\right) = 0.09$

Não há evidência suficiente para rejeitar H0, teste inconclusivo.
Não se pode afirmar que a média seja superior a 770€.

Decisão tomada com confiança moderada (9%) !

Teste à Diferença de Valores Esperados (amostras emparelhadas)

Exemplos de amostras emparelhadas:

- Idades do marido e da mulher de seis casais seleccionados ao acaso

  {(31, 27), (82, 76), (21 ,22), (27, 27), (53, 50), (67,62)}

  O marido de uma senhora de 80 anos terá maior probabilidade de ser idoso (amostras não são independentes)

- Classificações a uma dada disciplina dos alunos do 10º ano no 1º e no 3º período.

jlborges@fe.up.pt

amostras emparelhadas

analisar o comportamento da amostra univariada constituída pela diferença entre pares de observações

Note que as amostras **não são independentes** e a distribuição da estatística de teste é difícil de caracterizar.

{(31, 27), (82, 76), (21 ,22), (27, 27), (53, 50), (67,62)}

$$\Delta = \{31-27=4;\quad 82-76=6;\quad 21-22=-1;\quad 27-27=0;\quad 53-50=3;\quad 67-62=5\}$$

$$\overline{\Delta} = 2.83 \quad S_{\Delta}^{2} = 2.787^{2}$$

Para verificar se o valor esperado das **idades dos maridos** (A) é, ou não, **superior** ao valor esperado das **idades das mulheres** (B):

$$\Delta = x_A - x_B$$

$$\Delta = \{31-27=4;\quad 82-76=6;\quad 21-22=-1;\quad 27-27=0;\quad 53-50=3;\quad 67-62=5\}$$

$$\bar{\Delta} = 2.83 \quad S_\Delta^2 = 2.787^2$$

$$H_0: \quad \mu_\Delta = 0$$
$$H_1: \quad \mu_\Delta > 0$$

$$ET = \frac{\bar{\Delta} - \Delta_0}{S_\Delta / \sqrt{N}} = \frac{2.83 - 0}{2.787/\sqrt{6}} = 2.49 > t_{(6-1)}(0.05) = 2.015$$

Valor de Prova: 2.76%

Conclui-se então que, em média, os maridos têm idade superior à das mulheres.

Relação entre Testes de Hipóteses e Intervalos de Confiança

Uma HIPÓTESE NULA (**H0: $\mu = \mu 1$**) **pode ser rejeitada** a um nível de significância α se, e só se, o

intervalo de confiança de μ a $(1-\alpha).100\%$

não incluir o valor $\mu 1$

O intervalo de confiança deverá ser compatível com a natureza de H_1:

se o teste for bilateral o IC deverá ser bilateral

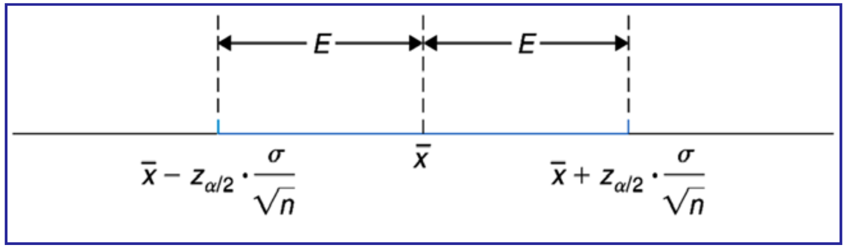
e

se o teste for unilateral o IC deverá ser unilateral

Exemplo (teste bilateral):

Para umapopulação Normal com $\sigma^2=1$ e uma amostra de N=25, obteve-se $\bar{x}=1.4$
Verifique a hipótese H_0:μ=1 com H_1: $\mu \neq 1$.

$$H_1: \mu \neq 1 \quad \rightarrow \quad \text{IC bilateral}$$



$$\bar{X} \pm z(\alpha/2) \cdot \frac{\sigma}{\sqrt{N}}$$

$$\alpha = 5\% \quad \Rightarrow \quad z(0.025) = 1.96$$

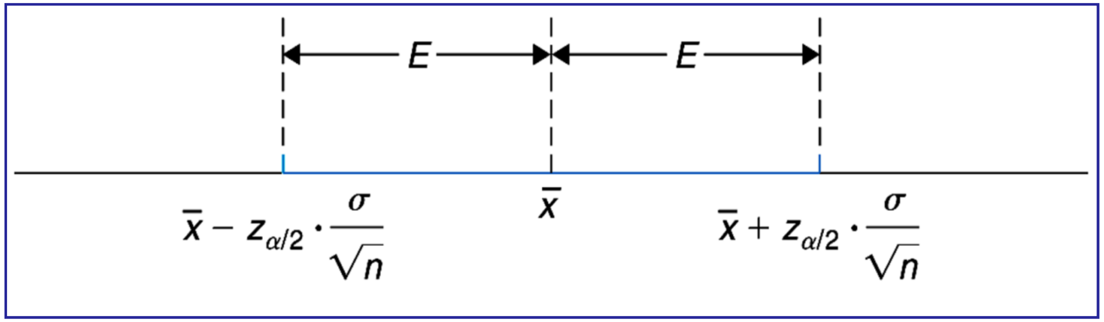
$$\bar{x} \pm z(\alpha/2) \cdot \frac{\sigma}{\sqrt{N}} = 1.4 \pm 1.96 \cdot \frac{1}{\sqrt{25}} \equiv \left[1.01,\ 1.79\right]$$

Como 1 não pertence ao intervalo a hipótese nula deverá ser rejeitada.

Exemplo (cont.):

Qual o valor de prova?

O valor de prova corresponde ao menor grau de confiança que permite que o intervalo inclua o valor 1, logo:



$$\bar{x} - z(\alpha/2) \cdot \frac{\sigma}{\sqrt{N}} = 1 \qquad \Rightarrow z(\alpha/2) = (1.4 - 1) \cdot \sqrt{25} = 2.0$$

das tabelas vem

$$z(\alpha/2) = 2.0 \quad \rightarrow \quad \alpha/2 = 0.0228$$

sendo o valor de prova

$$\alpha = 2 \cdot 0.0228 = 4.56\%$$

Exemplo (teste unilateral):

Para duas amostras de N_A=16 e N_B=21, retiradas de duas populações Normais, obtiveram-se as seguintes estimativas

$$S_A^2 = 4.60 \qquad S_B^2 = 1.12$$

Teste se a variância de A é significativamente superior à variância de B.

$$H_0 : \sigma_A^2 = \sigma_B^2 \quad \equiv \quad H_0 : \frac{\sigma_A^2}{\sigma_B^2} = 1$$

$$H_1 : \sigma_A^2 > \sigma_B^2 \quad \equiv \quad H_1 : \frac{\sigma_A^2}{\sigma_B^2} > 1$$

Sendo o I. C.

$$\left[\frac{1}{F_{N_A-1, N_B-1(\alpha)}} \cdot \frac{S_A^2}{S_B^2}, \infty \right[= \left[\frac{1}{2.20} \cdot \frac{4.60}{1.12}, \infty \right[$$

$$= [1.87, \infty[$$

$$\Rightarrow \text{ Rejeitar } H_0$$

Valor de prova:

$$\frac{1}{F_{N_A-1, N_B-1(\alpha)}} \cdot \frac{S_A^2}{S_B^2} = 1 \Leftrightarrow F_{N_A-1, N_B-1(\alpha)} = \frac{4.60}{1.12} = 4.11 \equiv \alpha = 0.19\%$$

Erro do Tipo II

O nível de significância do teste, α, representa a probabilidade de se rejeitar H_0 quando esta hipótese é verdadeira. **Erro do Tipo I**.

Existe ainda a possibilidade de cometer um outro tipo de erro que corresponde a não rejeitar H0 quando esta é falsa. **Erro Tipo II (β).**

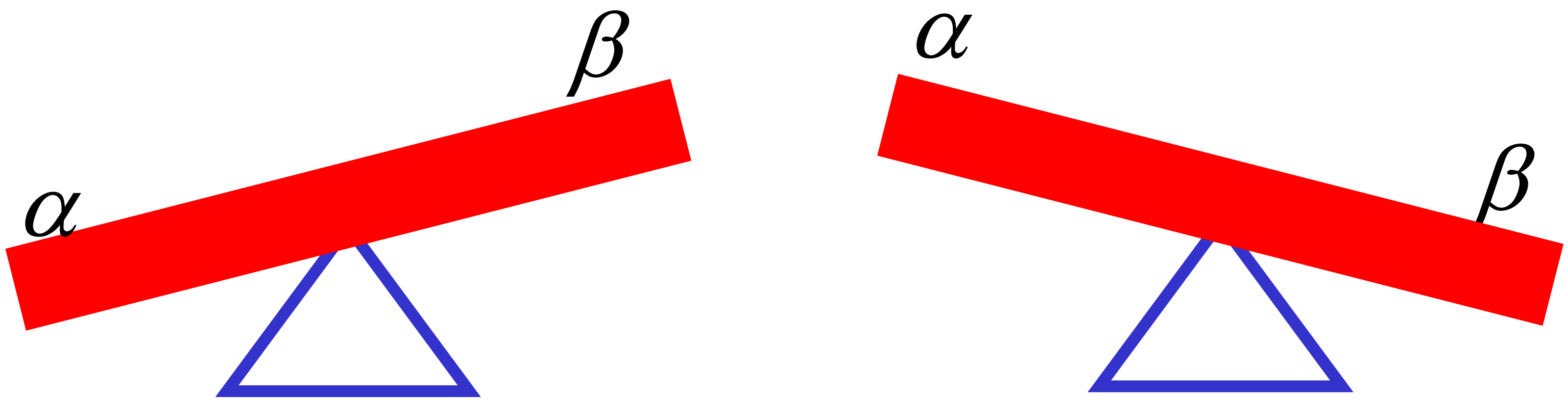
	H_0 Verdadeira	H_0 Falsa
H_0 Rejeitada	Erro Tipo I (α)	Decisão Correcta (1-β)
H_0 Não Rejeitada	Decisão Correcta (1-α)	Erro Tipo II (β)

jlborges@fe.up.pt

Potência do Teste

A **potência do teste** traduz a probabilidade de rejeitar H_0 quando esta hipótese é falsa e designa-se **1-** β

O desejável era que α e β fossem ambos pequenos, mas quando α diminui β aumenta, diminuindo consequentemente a potência do teste

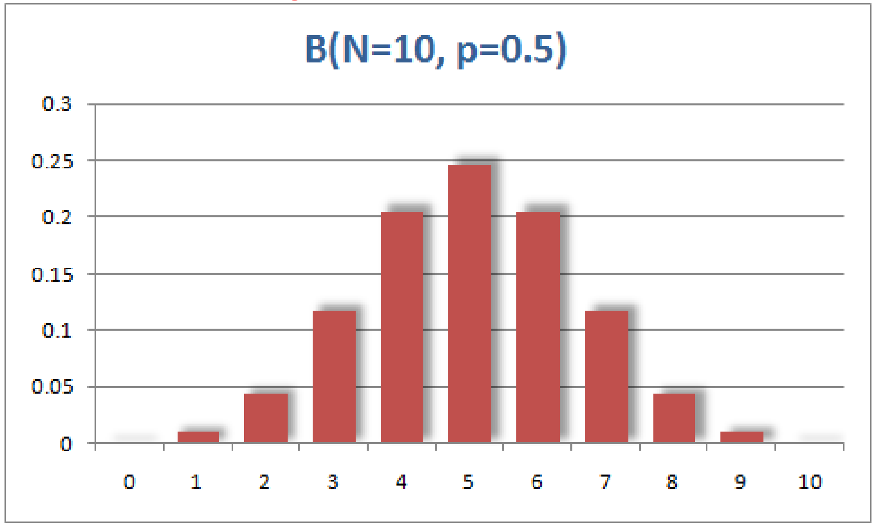


jlborges@fe.up.pt

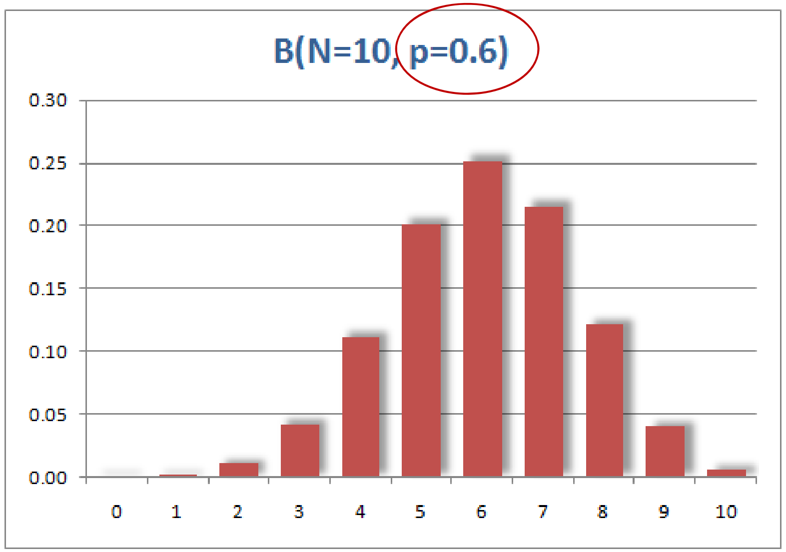
Teste à moeda. Lançar 10 vezes e verificar probabilidade do resultado obtido

H0: p = 0.5 H1: p <> 0.5

H0:p=0.5 Verdadeira



H0 Falsa



Erro Tipo 1: probabilidade de a moeda ser 'boa' e obter 9 ou mais caras.

Erro Tipo 2: probabilidade de a moeda ser 'má' e obter 8 ou menos caras.

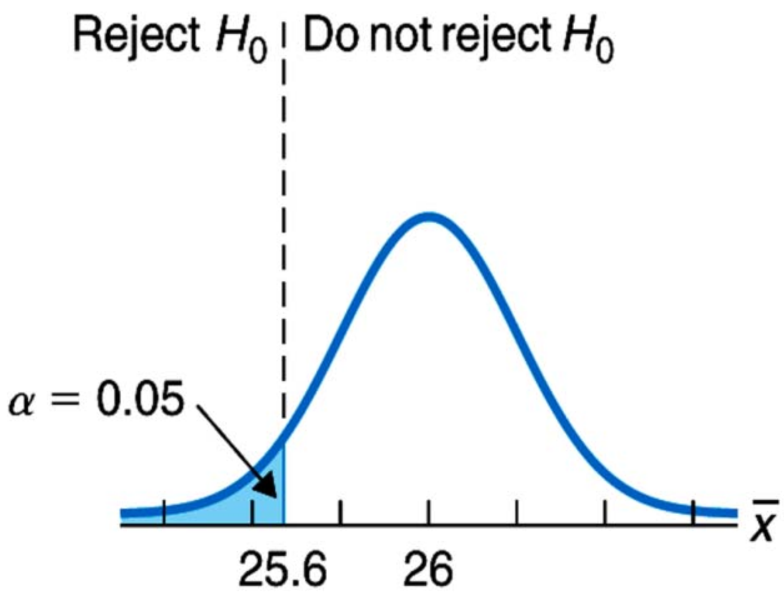
Este teste é muito pouco potente (ver excel).

Exemplo (cont.)

verificar a conjectura acerca do tempo médio de transferência ($\mu < 26$)

$N = 50 \quad \bar{x} = 25.51 \quad \sigma^2 = 1.84^2$

$H_0 : \mu \geq 26$ ou $\mu = 26$

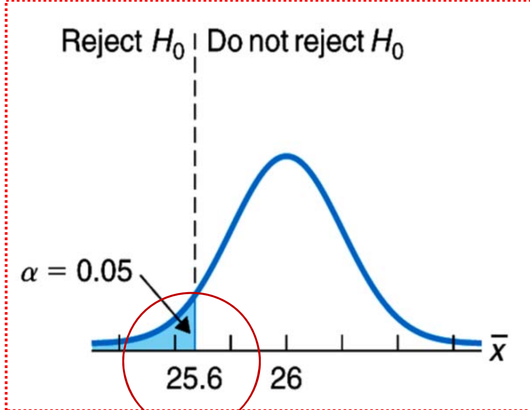


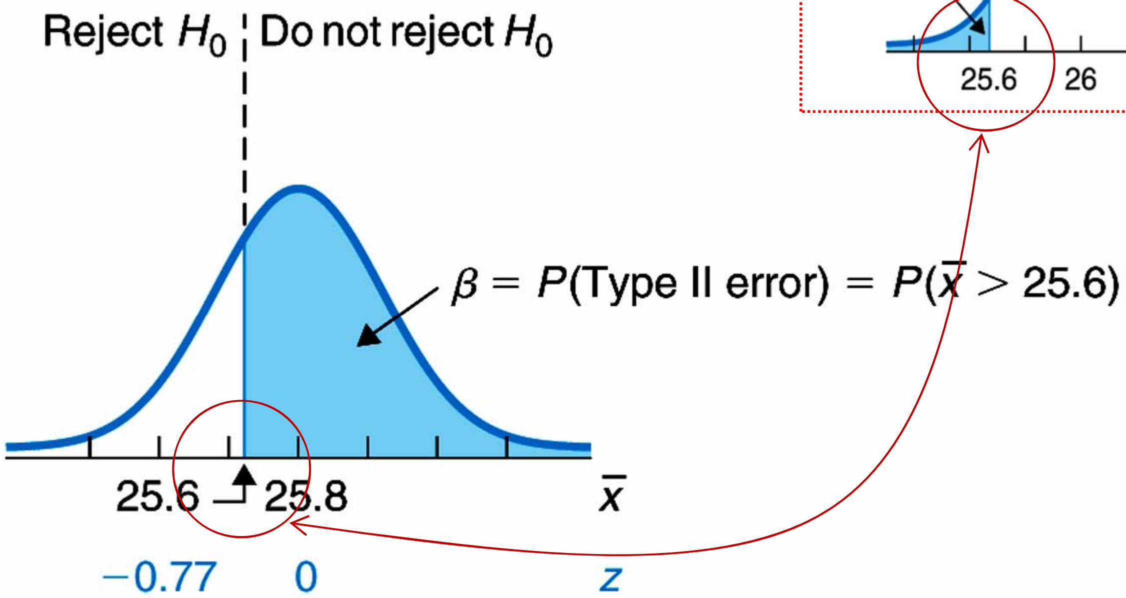
Se o valor amostral for inferior ao Valor Crítico a hipótese nula deverá ser rejeitada.

Exemplo (cont.)

H0 é falsa

Cálculo de β se μ**=25.8** e α=5%:





z-score computation:

$$\bar{x} = 25.6 \longrightarrow z = \frac{25.6 - 25.8}{0.26} = -0.77$$

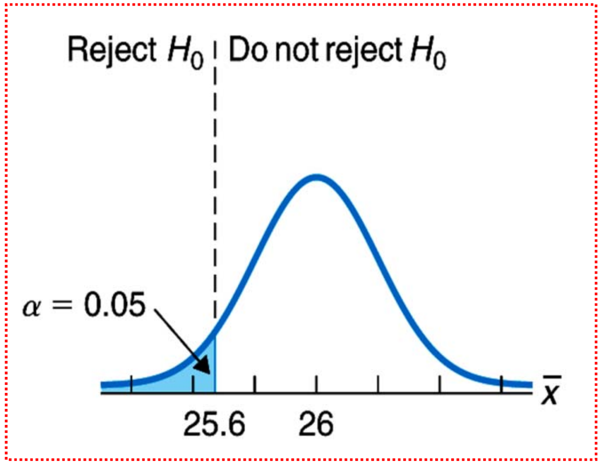
Area to the left of *z*:

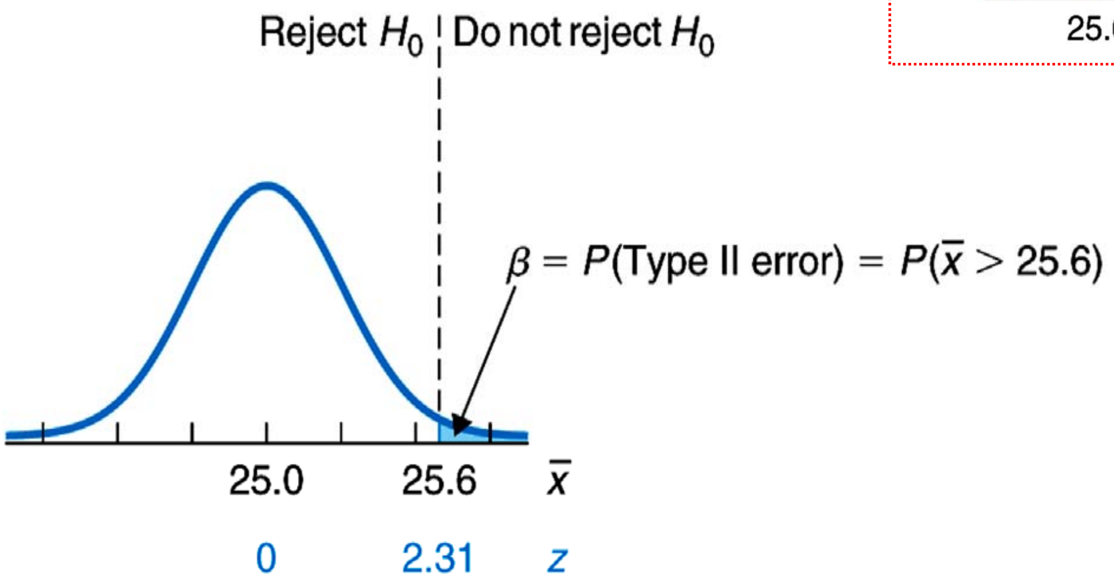
0.2206 **Rejeitar H0 (decisão correcta)**

Shaded area = 1 − 0.2206 = 0.7794 **Não Rejeitar H0 (decisão incorrecta)**

Exemplo (cont.)

Cálculo de β se **μ=25.0** e α=5%:





z-score computation:

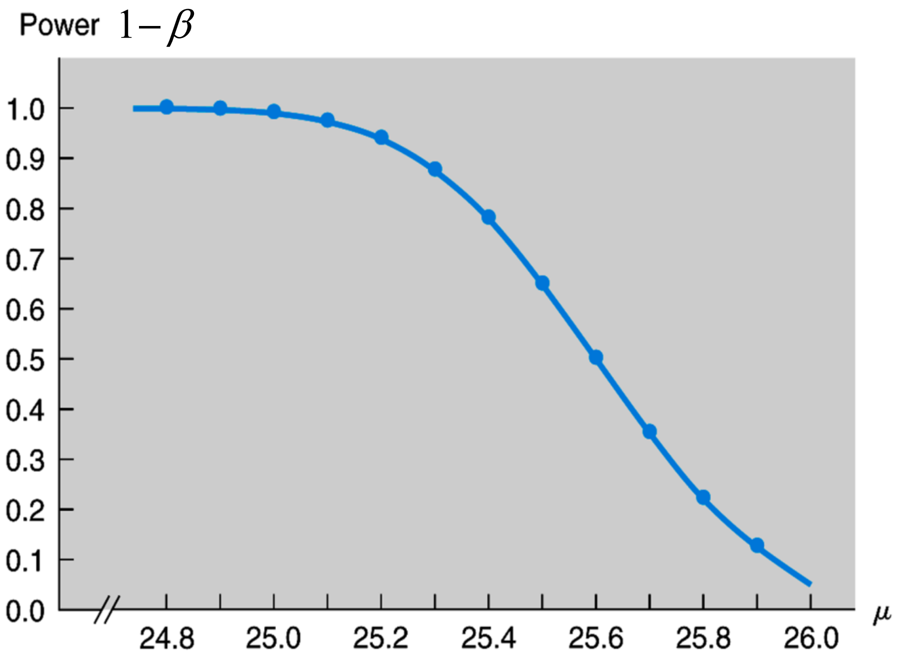
$$\bar{x} = 25.6 \longrightarrow z = \frac{25.6 - 25.0}{0.26} = 2.31$$

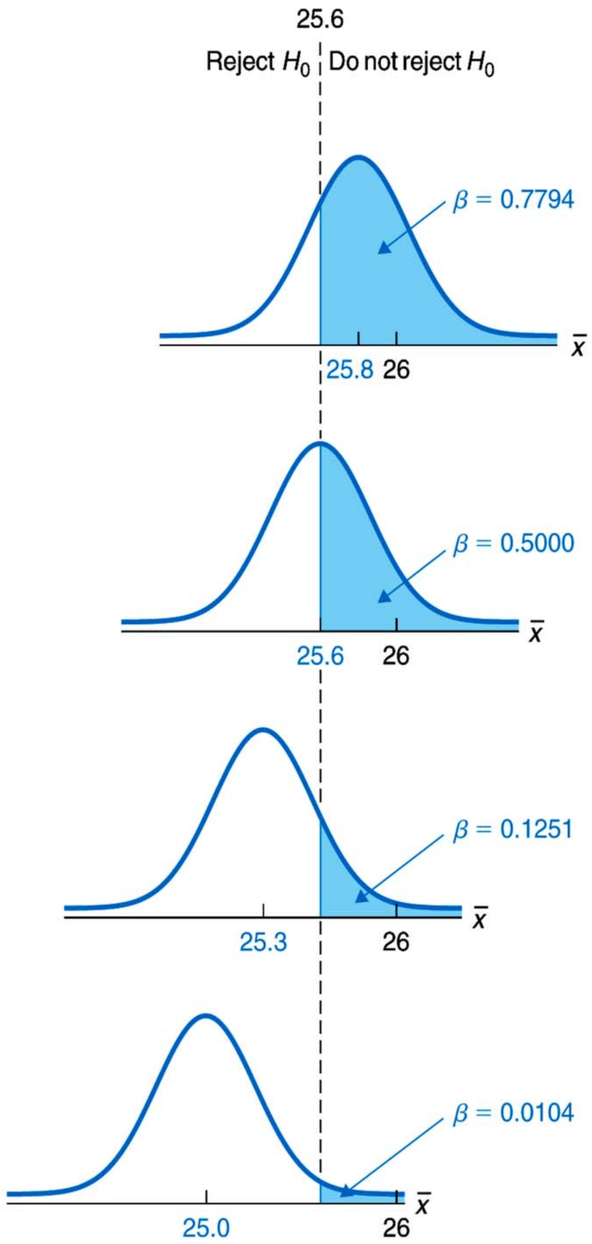
Area to the left of *z*:

0.9896

Shaded area = 1 − 0.9896 = 0.0104

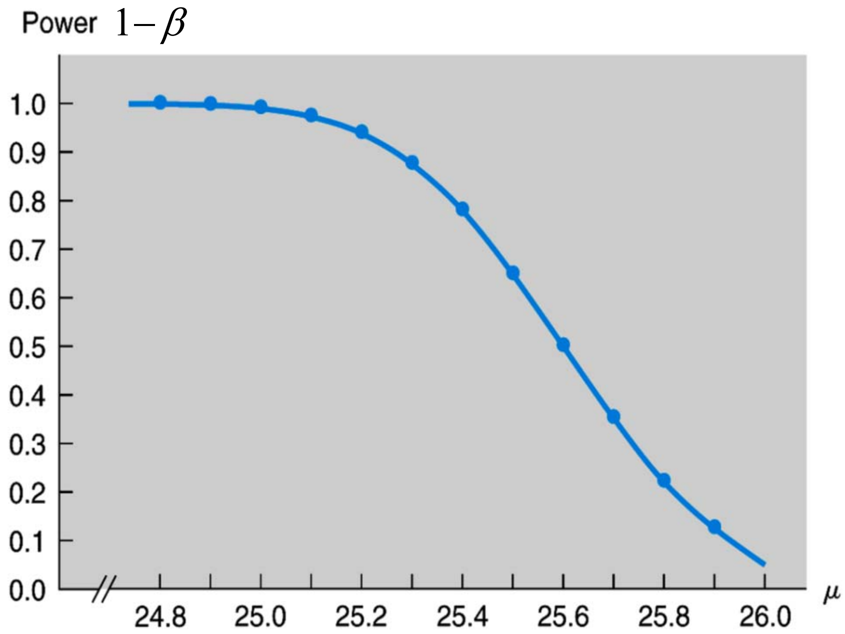
Valor de β para μ=25.8, 25.6, 25.3 e 25 (α=5%:)





Quanto maior for μ_1- μ_0 maior será a potência do teste 1-b.

Só faz sentido calcular a potência de teste porque não sabemos o valor verdadeiro de μ



Imagine que o chefe disse ao responsável pela rede que ia recolher uma amostra de tempos de transferência para verificar se este é inferior a 26seg.

Se não se rejeitar a hipótese μ >= 26, o responsável da rede será despromovido.

Para qual valor de μ ele deve apontar no que diz respeito à configuração da rede?

A análise da potência de teste permite analisar cenários alternativos.

Para <u>diminuir</u> α <u>e</u> β <u>simultaneamente</u> torna-se necessário aumentar a dimensão da amostra, provocando a redução da variância da distribuição da estatística de teste

Exemplo

Uma amostra aleatória de 80 associados de um clube automóvel revelou que o montante gasto em prémios de seguro automóvel tem um valor médio de 800 Euros. O desvio padrão da variável é conhecido e toma o valor de 200 Euros

i) Verifique se é razoável afirmar que o valor médio gasto pelos sócios do clube em prémios de seguros é superior a 770 Euros.

ii) Se de facto o valor médio gasto em prémios de seguros for de 780 Euros qual o valor da potência do teste efectuado em (i) ?

Note que nós não sabemos qual o valor de μ e apenas pretendemos estudar o comportamento do teste de hipóteses no contexto de um cenário hipotético.

Resolução (cont.):

ii) se $\mu = 780$ que valor toma $1-\beta$?

$x_c = ?$

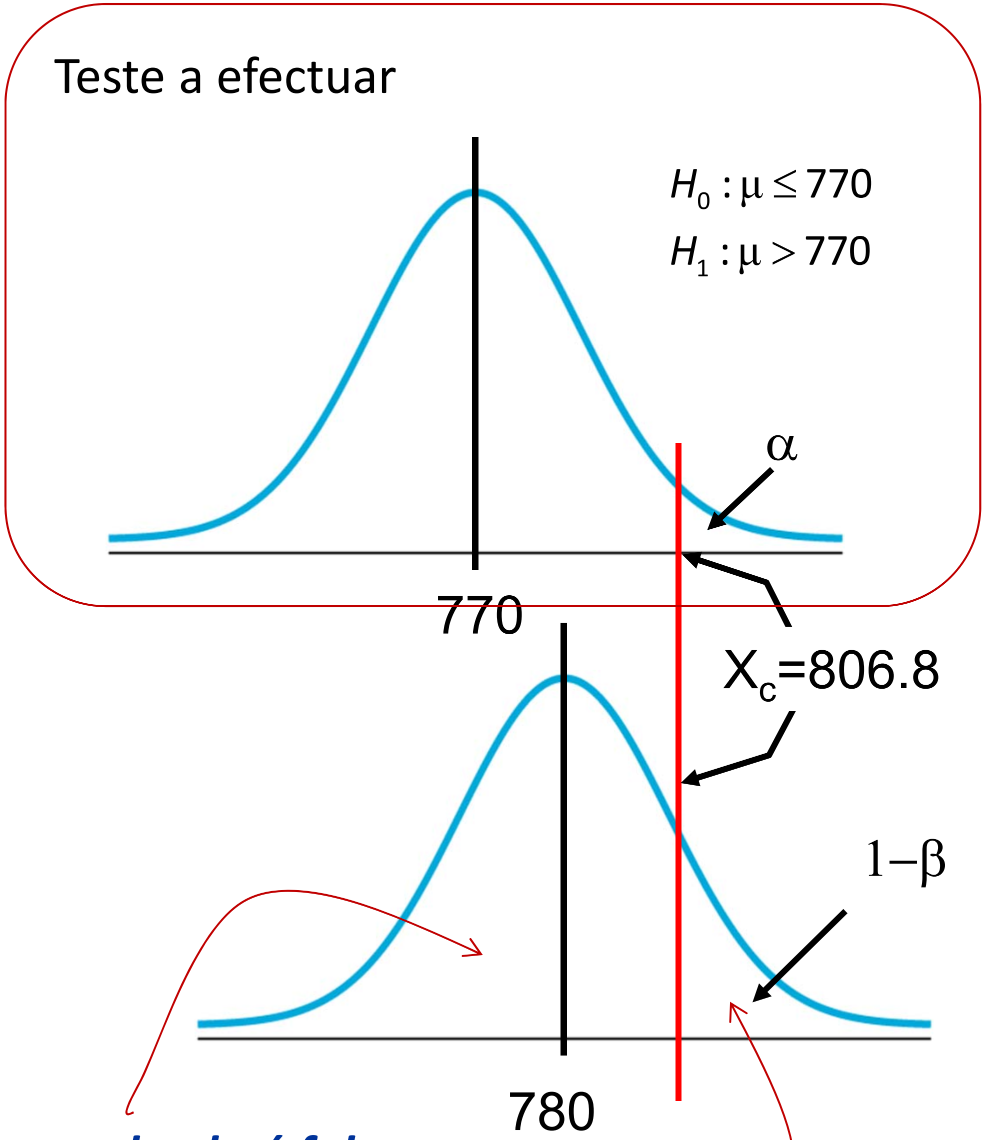
$$\frac{x_c - \mu}{\sigma / \sqrt{N}} = z(0.05) \Leftrightarrow$$

$$x_c = 1.645 \cdot 200 / \sqrt{80} + 770 = 806.78$$

$$P(1-\beta) = P\left(z > \frac{806.78 - 780}{200/\sqrt{80}}\right)$$

$$= P(z > 1.198)$$

$$= 11.5\%$$



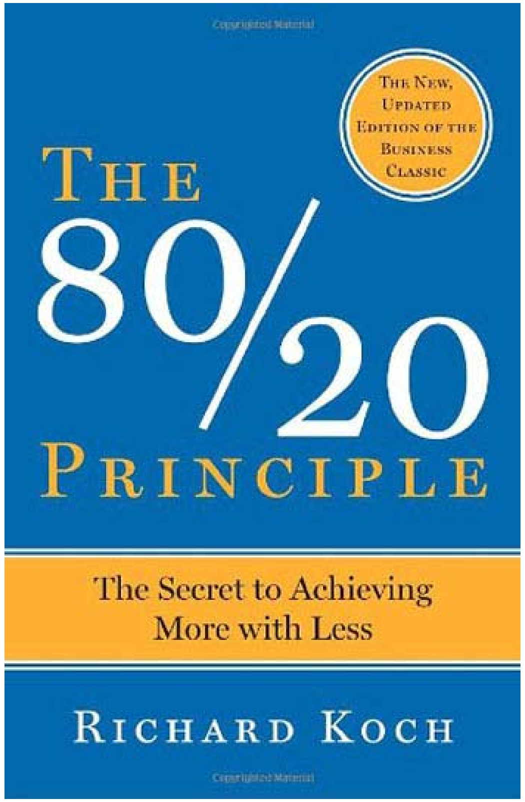
Probabilidade de NÂO rejeitar H0 quando ela é falsa.

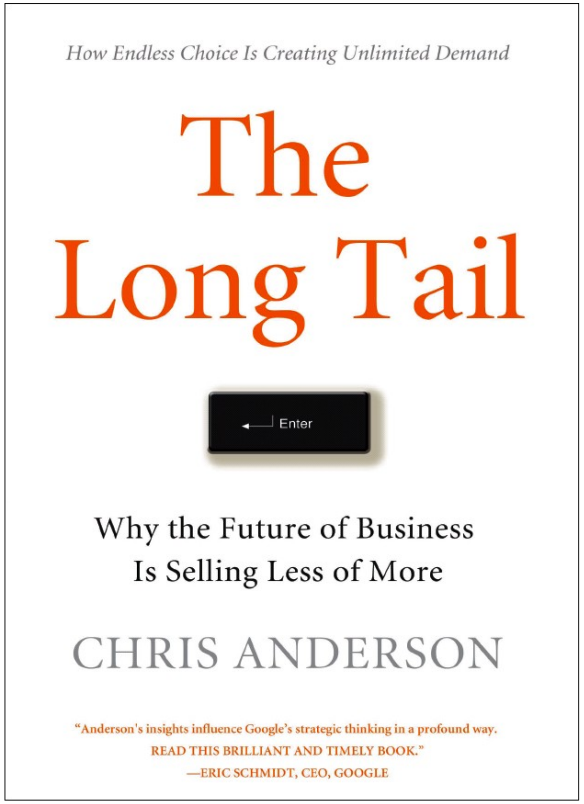
Probabilidade de rejeitar H0 quando ela é falsa.

Resultados de Aprendizagem

- Saber especificar e realizar testes de hipóteses bi-laterais ou uni-laterais para:
  - a média
  - variância
  - proporção binomial
- e para
  - a diferença das médias (com amostras emparelhadas ou não)
  - a razão das variâncias
  - a diferença de proporções binomiais
- Saber calcular o valor de prova, o erro do tipo II e a potência de teste para qualquer teste de hipóteses

The Pareto Principle and The Long Tail





What is the 80/20 (Pareto) Principle?

- The 80/20 Principle asserts that a minority of causes, inputs, or effort usually lead to a majority of the results, outputs, or rewards.
- A typical pattern will show that 80% of outputs result from 20% of inputs
  - 20% of products usually account for about 80% of sales value
  - 20% of criminals account for 80% of the value of all crime
  - In the home, 20% of your carpets are likely to get 80% of the wear
  - 20% of the population enjoyed 80% of the wealth
  - ???? 20% of the study will lead to 80% of your mark ????

Vilfredo Pareto

Economist

Vilfredo Federico Damaso Pareto was an Italian engineer, sociologist, economist, political scientist and philosopher. Wikipedia

Born: July 15, 1848, Paris

Died: August 19, 1923, Céligny

Books: The transformation of democracy

Education: HEC Lausanne, Polytechnic University of Turin

Applications of the Pareto Principle

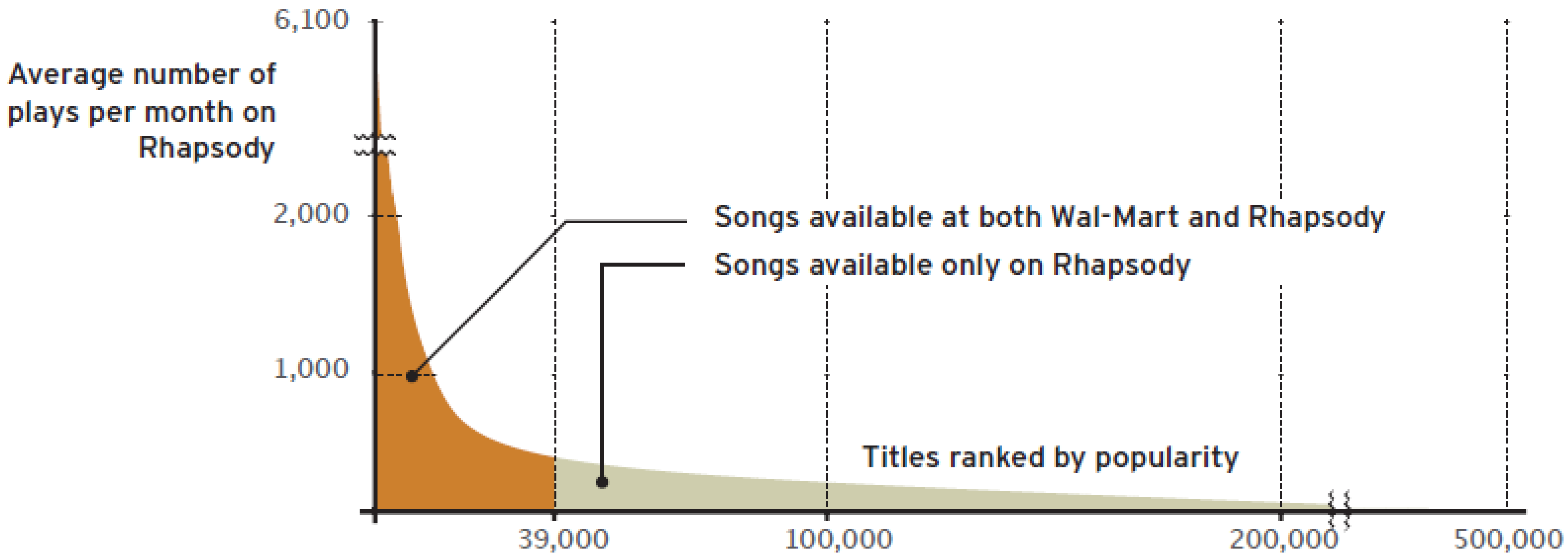
- IBM was one of the earliest and most successful corporations to spot and use the 80/20 Principle.
- In 1963, IBM discovered that about 80% of a computer's time is spent executing about 20% of the operating code.
- The company immediately rewrote its operating software to make the most-used 20% very accessible and user friendly, thus making IBM computers more efficient and faster than competitors' machines for the majority of applications.

Applications of the Pareto Principle

- 80/20 => retailers will carry only content that can generate sufficient demand to earn its keep on the shelves
  - music
  - dvds
  - books
  - …
- We've been suffering the tyranny of lowest-common-denominator fare, subjected to blockbusters and manufactured pop. Why?
- Economics. Many of our assumptions about popular taste are actually artifacts of poor supply-and-demand matching - a market response to inefficient distribution.

The Long Tail

- Online sellers enable unlimited selection.
- Forget squeezing millions from a few megahits at the top of the charts. The future of entertainment is in the millions of niche markets at the shallow end of the bitstream.

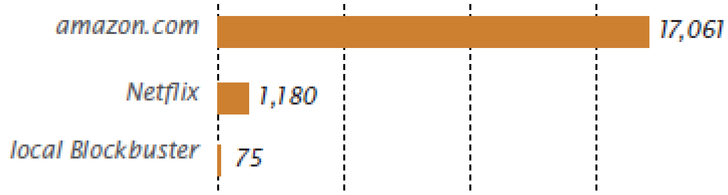


- Make everithing available.

THE DOCUMENTARY NICHE GETS RICHER [2]

More than 40,000 documentaries have been released, according to the Internet Movie Database. Of those, Amazon.com carries 40 percent, Netflix stocks 3 percent, and the average Blockbuster just .2 percent.

DOCUMENTARIES AVAILABLE



- Cut the price in half. Now lower it.

THE REAL COST *of* MUSIC [3]

Online music services don't incur packaging, distribution, and retail fees — and they should charge accordingly.

CREATION COSTS

Artist	$1.50
Marketing and Profit	$5.00
Publishing	$0.96
	$7.46

+

PRODUCTION COSTS

Packaging	$0.75
Distribution	$2.00
Retail markup	$5.00
	$7.76

$15.21/CD

Divided by 12 tracks
=
$0.62/track
+
17¢ online delivery cost

$0.79/song

- Help me find it.

"IF YOU LIKE BRITNEY, YOU'LL LOVE...[4]"

Just as lower prices can entice consumers down the Long Tail, recommendation engines drive them to obscure content they might not find otherwise.