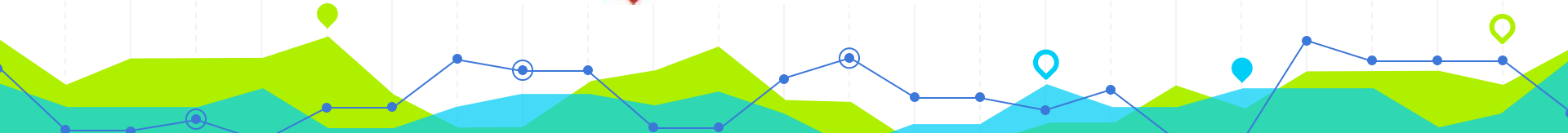




Lisbon School
of Economics
& Management

Universidade de Lisboa



Estatística II

Licenciatura em Gestão do Desporto
2.º Ano/2.º Semestre
2025/2026

Aulas Teórico-Práticas N.ºs 19 e 20 (Semana 11)

Docente: Elisabete Fernandes

E-mail: efernandes@iseg.ulisboa.pt



<https://doity.com.br/estatistica-aplicada-a-nutricao>



<https://basiccode.com.br/produto/informatica-basica/>

Conteúdos Programáticos

Aulas Teórico-Práticas (Semanas 1 a 3)

- **Capítulo 1:** Revisões e Distribuições de Amostragem

Aulas Teórico-Práticas (Semanas 4 a 7)

- **Capítulo 2:** Estimação

Aulas Teórico-Práticas (Semanas 7 a 9)

- **Capítulo 3:** Testes de Hipóteses

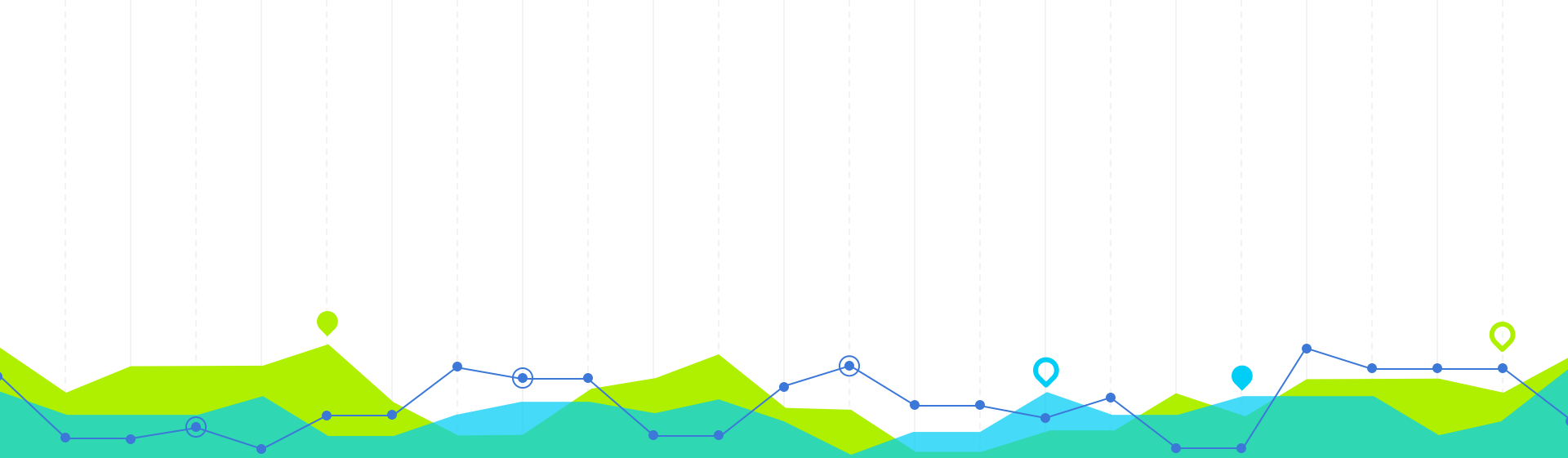
Aulas Teórico-Práticas (Semanas 10 a 13)

- **Capítulo 4:** Modelo de Regressão Linear Múltipla

Material didático: Exercícios do Livro Murteira et al (2015), Formulário e Tabelas Estatísticas

Bibliografia: B. Murteira, C. Silva Ribeiro, J. Andrade e Silva, C. Pimenta e F. Pimenta; *Introdução à Estatística*, 2ª ed., Escolar Editora, 2015.

<https://cas.iseg.ulisboa.pt>



Testes de Hipóteses para σ_1^2/σ_2^2

Hipóteses, Estatística de Teste e Decisão

1

Teste de Hipóteses para σ_1^2/σ_2^2 : Formulário

• POPULAÇÕES NORMAIS

Variância corrigida

Média	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1)$	$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
Diferença de médias	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \sim t(v)$	
	$T = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	onde v é o maior inteiro contido em r, $r = \frac{\left(\frac{S_1'^2}{m} + \frac{S_2'^2}{n}\right)^2}{\frac{1}{m-1} \left(\frac{S_1'^2}{m}\right)^2 + \frac{1}{n-1} \left(\frac{S_2'^2}{n}\right)^2}$	
Variância	$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$		
Relação de variâncias	$\frac{S_1'^2}{S_2'^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$		

Teste de Hipóteses para σ_1^2/σ_2^2

Considere que duas populações Normais, com médias μ_1 e μ_2 e desvios padrão σ_1 e σ_2 , das quais se extraíram aleatoriamente duas amostras independentes com dimensão n_1 e n_2 , respectivamente. Pretende-se realizar um teste de hipóteses para comparar as variâncias populacionais σ_1^2 e σ_2^2 , i. e., para o **quociente de variâncias** (σ_2^2/σ_1^2). A estatística de teste a utilizar é:

$$F = \frac{S_1^2}{S_2^2} \left(\frac{\sigma_2^2}{\sigma_1^2} \right)_0 \sim F_{n_1-1; n_2-1},$$

onde $(\sigma_2^2/\sigma_1^2)_0$ representa o valor que se assume para (σ_2^2/σ_1^2) em H_0 .

Teste de Hipóteses para σ_1^2/σ_2^2

Nesta secção considera-se apenas a situação em que $(\sigma_2^2/\sigma_1^2)_0 = 1$, o que equivale a comparar a igualdade das variâncias populacionais.

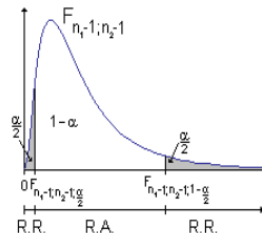
T. bilateral

Hipóteses a testar:

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ vs } H_1: \sigma_1^2 \neq \sigma_2^2 \Leftrightarrow$$

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \text{ vs } H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

Regiões críticas:



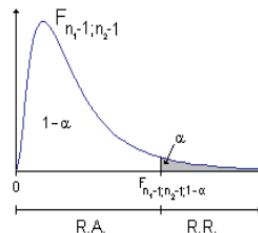
T. unilateral direito

Hipóteses a testar:

$$H_0: \sigma_1^2 \leq \sigma_2^2 \text{ vs } H_1: \sigma_1^2 > \sigma_2^2 \Leftrightarrow$$

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} \leq 1 \text{ vs } H_1: \frac{\sigma_1^2}{\sigma_2^2} > 1$$

Regiões críticas:



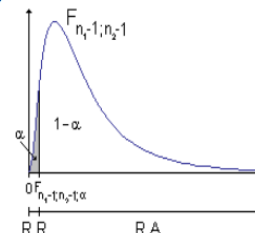
T. unilateral esquerdo

Hipóteses a testar:

$$H_0: \sigma_1^2 \geq \sigma_2^2 \text{ vs } H_1: \sigma_1^2 < \sigma_2^2 \Leftrightarrow$$

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} \geq 1 \text{ vs } H_1: \frac{\sigma_1^2}{\sigma_2^2} < 1$$

Regiões críticas:



Nos primeiros 6 meses de vida dois grupos aleatórios de crianças seguiram esquemas de alimentação diferentes: o grupo 1 seguiu o esquema A e o grupo 2 seguiu o esquema B. No quadro seguinte apresentam-se os ganhos em peso, em kg, dessas crianças.

Grupo 1	2,7	3,2	3,6	4,1	2,7	3,2	4,5	3,6	2,7
Grupo 2	4,1	4,5	3,6	2,7	3,6	3,2	4,1		

Sabe-se que as crianças dos dois grupos tinham, ao nascer, aproximadamente pesos iguais.

Admita que as distribuições dos pesos seguem a distribuição Normal.

- Com base num teste de hipóteses, ao nível de significância de 5%, pode concluir que a variabilidade é igual nos dois grupos?
- Sem efectuar cálculos diga, justificando, qual a decisão que tomava no âmbito da alínea b), se considerasse um nível de significância de 1%?
- Para o teste da alínea a) calcule o respectivo valor p e interprete.

[ProbabilidadesEstatistica_2019\(uevora.pt\)](http://ProbabilidadesEstatistica_2019(uevora.pt))



Exercício (a)

Sejam:

- X_1 a v.a. que representa o ganho em peso, em kg, das crianças alimentadas segundo o esquema A,
- X_2 a v.a. que representa o ganho em peso, em kg, das crianças alimentadas segundo o esquema B,

Com $X_1 \sim N(\mu_1 = ?; \sigma_1 = ?)$ e $X_2 \sim N(\mu_2 = ?; \sigma_2 = ?)$.

$$n_1 = 9, \quad \bar{x}_1 = 3,3667 \quad \text{e} \quad s_1^2 = 0,4150,$$

$$n_2 = 7, \quad \bar{x}_2 = 3,6857 \quad \text{e} \quad s_2^2 = 0,3714.$$

a) $\alpha = 5\%$, $\sigma_1^2 = \sigma_2^2$?

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ vs } H_1: \sigma_1^2 \neq \sigma_2^2$$

$$\Leftrightarrow H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \text{ vs } H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \text{ (teste bilateral).}$$

Exercício (a)

Estatística de teste:

$$F = \frac{S_1^2}{S_2^2} \left(\frac{\sigma_2^2}{\sigma_1^2} \right)_0 \sim F_{n_1-1; n_2-1=8; 6}.$$

$$f_{obs} = \frac{0,4150}{0,3714} \times 1 = 1,1173.$$

Pela tabela, $f_{n_1-1; n_2-1; \frac{\alpha}{2}} = f_{8; 6; 0,25} = \frac{1}{f_{6; 8; 0,975}} = \frac{1}{4,65} = 0,215$ e $f_{n_1-1; n_2-1; 1-\frac{\alpha}{2}} = f_{8; 6; 0,975} = 5,6$.

Logo, $R.A. :]0,215; 5,6[$ e $R.R. : [0; 0,215] \cup [5,6; +\infty[$

Como $f_{obs} \in R.A.$ não rejeitar H_0 . Portanto, ao nível de significância de 5%, não existe evidência estatística de que a variabilidade nos pesos seja significativamente diferente nos dois grupos.

Exercícios (b) e (c)

- b) Se $\alpha = 1\%$ a decisão tomada no teste anterior era a mesma, ou seja, não rejeitar H_0 . Esta situação é originada pelo facto de quando se diminui o nível de significância também se diminui a $R.R.$ e consequentemente a $R.A.$ aumenta. Portanto, se quando $\alpha = 5\%$ f_{obs} está na $R.A.$ então $\alpha = 1\%$ a situação mantém-se.
- c) valor $p = 2 \times \min\{P(F \leq f_{obs}); P(F \geq f_{obs})\}$
 $= 2 \times \min\{P(F \leq 1,1173); P(F \geq 1,1173)\}$
 $= 2 \times \min\{0,5409; 0,4591\}$
 $= 2 \times 0,4591 = 0,9181.$

A hipótese $H_0: \sigma_1^2 = \sigma_2^2$ é rejeitada para níveis de significância superiores ou iguais a 91,81%, indicando que não existe evidência de que as variâncias sejam diferentes.

34. Dois programas de alimentação de gado bovino são comparados. A variável aleatória X representa o aumento de peso (em kg) de um animal alimentado segundo o programa 1, durante um mês, enquanto Y traduz o aumento de peso (em kg) de um animal alimentado segundo o programa 2, durante igual período de tempo. Sabe-se que $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$, e que as variáveis são independentes. Um grupo de 8 animais foi submetido ao primeiro programa durante um mês, tendo-se obtido

$$\sum_{i=1}^8 x_i = 416 \text{ e } \sum_{i=1}^8 x_i^2 = 21807.$$

Outro grupo de 10 animais foi submetido ao segundo programa durante um mês, tendo-se obtido

$$\sum_{i=1}^{10} y_i = 468 \text{ e } \sum_{i=1}^{10} y_i^2 = 22172.$$

- a) Teste, ao nível de 0.05, a igualdade entre as variâncias.
- b) Teste, ao nível de 0.05, a igualdade entre os valores médios supondo variâncias iguais.
- c) Teste, ao nível de 0.05, a igualdade entre os valores médios supondo variâncias diferentes.
- d) Qual dos procedimentos adoptados nas alíneas anteriores, para testar a igualdade de médias, lhe parece mais adequado?



Exercício 34

$$X \sim N(\mu_1, \sigma_1^2)$$

$$n_1 = 8$$

$$\sum_{i=1}^8 x_i = 416$$

$$\sum_{i=1}^8 x_i^2 = 21807$$

$$Y \sim N(\mu_2, \sigma_2^2)$$

$$n_2 = 10$$

$$\sum_{i=1}^{10} y_i = 468$$

$$\sum_{i=1}^{10} y_i^2 = 22172$$

Exercício 34 a)

$$a) \quad \alpha = 0.05$$

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$$\text{sob } H_0, F = \frac{S_1'^2}{S_2'^2} \sim F(8-1, 10-1) = F(7, 9)$$

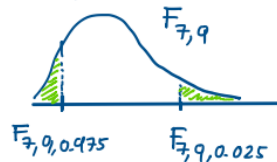
$$s_1'^2 = \sum_{i=1}^8 x_i^2 / 8 - \bar{x}^2 = \frac{21807}{8} - \left(\frac{416}{8}\right)^2 = 21.875$$

$$s_1'^2 = \frac{m_1}{m_1 - 1} s_1^2 = \frac{8}{7} \times 21.875 = 25$$

$$s_2'^2 = \sum_{i=1}^{10} y_i^2 / 10 - \bar{y}^2 = \frac{22172}{10} - \left(\frac{468}{10}\right)^2 = 26.96$$

$$s_2'^2 = \frac{m_2}{m_2 - 1} s_2^2 = \frac{10}{9} \times 26.96 = 29.9(5)$$

$$f_{\text{obs}} = \frac{s_1'^2}{s_2'^2} = \frac{25}{29.9(5)} \simeq 0.8346$$



Exercício 34 a)

$$\begin{aligned}W_F &= \} f_{obs}: f_{obs} < F_{7,9,0.975} \vee f_{obs} > F_{7,9,0.025} \{ \\ &= \} f_{obs}: f_{obs} < 0.21 \vee f_{obs} > 4.82 \{ \end{aligned}$$

$$F_{7,9,0.025} = 4.82 \quad (\text{tabela})$$

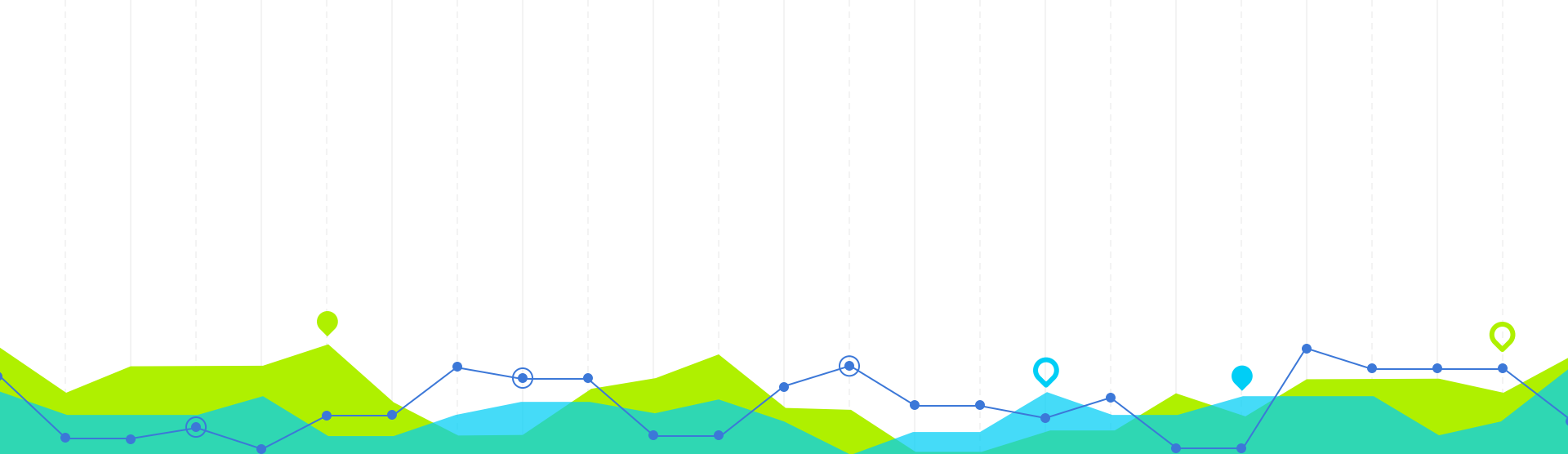
$$P(F < F_{7,9,0.975}) = 0.025 \quad (\Rightarrow)$$

$$(\Rightarrow) P\left(\frac{1}{F} > \frac{1}{F_{7,9,0.975}}\right) = 0.025 \quad \frac{1}{F} \sim F(9,7)$$

$$\frac{1}{F_{7,9,0.975}} = F_{9,7,0.025} \quad (\Rightarrow)$$

$$(\Rightarrow) F_{7,9,0.975} = \frac{1}{F_{9,7,0.025}} = \frac{1}{4.82} \approx 0.21$$

$f_{obs} \notin W$ logo não se rejeita H_0 .



Testes de Hipóteses para p

Hipóteses, Estatística de Teste e Decisão

2

Teste de Hipóteses para p: Formulário

População de Bernoulli

Proporção	$\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} N(0,1)$
Diferença de proporções	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{m} + \frac{\theta_2(1-\theta_2)}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{m} + \frac{\bar{X}_2(1-\bar{X}_2)}{n}}} \stackrel{a}{\sim} N(0,1)$
Igualdade de proporções	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \hat{\theta}(1-\hat{\theta})}} \stackrel{a}{\sim} N(0,1)$ onde $\hat{\theta} = \frac{m\bar{X}_1 + n\bar{X}_2}{m+n}$	

Estimador do
Parâmetro p

Teste de Hipóteses

Intervalo de Confiança

$$\frac{P^* - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \stackrel{apr}{\sim} N(0, 1)$$

$$\left[\bar{P} - z_{0,95} \sqrt{\frac{\bar{P}(1-\bar{P})}{n}}; \bar{P} + z_{0,95} \sqrt{\frac{\bar{P}(1-\bar{P})}{n}} \right]$$

Exercício:

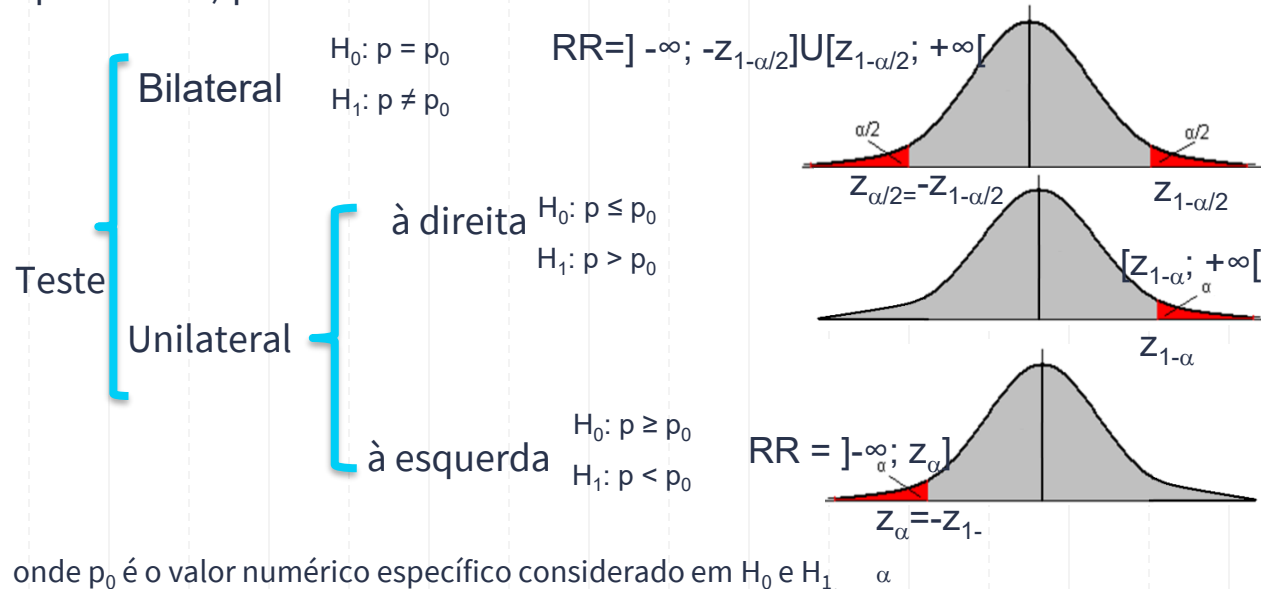
Considere-se uma amostra correspondente a 699 portugueses, sendo 205 fumadores.

Realize um Teste de Hipóteses para testar se a percentagem de fumadores portugueses é diferente de 40% para o nível de significância $\alpha = 1\%$.



Tipos de Testes de Hipóteses para uma Proporção Populacional

Um teste de hipóteses paramétrico para o parâmetro p (proporção populacional) pode ser:



Nota: Se $n \times p^* \geq 5$ e $n \times q^* \geq 5$ (corresponde a n grande), então a distribuição da estatística de teste é aproximadamente normal pelo Teorema de Limite Central (TLC).

Teste de Hipóteses para uma Proporção Populacional

$$n \times p^* = 204,8 \geq 5 \text{ e } n \times q^* = 494,2 \geq 5$$

Hipóteses

$$H_0: p = 0,4 \text{ versus } H_1: p \neq 0,4$$

Estatística de teste

$$\frac{P^* - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \overset{\text{apr}}{\sim} N(0, 1)$$

Valor observado da estatística de teste (VOE) $z_0 = z_{\text{obs}} = -5,775$

Teste bilateral

Dados:

$$n = 699 > 30$$

$$p^* = 205/699$$

$$p_0 = 0,40$$

$$q_0 = 0,60$$

$$\alpha = 0,01$$

$$z_{1-\alpha/2} = z_{0,995} = 2,576$$

Tabela da Distribuição Normal

Tabela da Distribuição Normal

Regra: $z_0 \in RR \Rightarrow$ Rejeita-se H_0

Região de rejeição ou crítica:

$$RR =] -\infty; -z_{0,995}] \cup [z_{0,995}; +\infty[=] -\infty; -2,576] \cup [2,576; +\infty[$$

Decisão

Pela região de rejeição: $z_0 = -5,775 \in RR =] -\infty; -2,576] \cup [2,576; +\infty[$

Pelo valor-p: Valor-p = $2 \times P(Z \geq |-5,775|) = 2 \times [1 - P(Z \leq 5,775)] \sim 2 \times (1 - 0,0005) = 0,001 < 0,01$

Rejeita-se H_0 para $\alpha = 1\%$. Existe evidência estatística para afirmar que, para $\alpha=1\%$, a proporção de indivíduos na pop. com a característica em estudo é diferente de 40%.

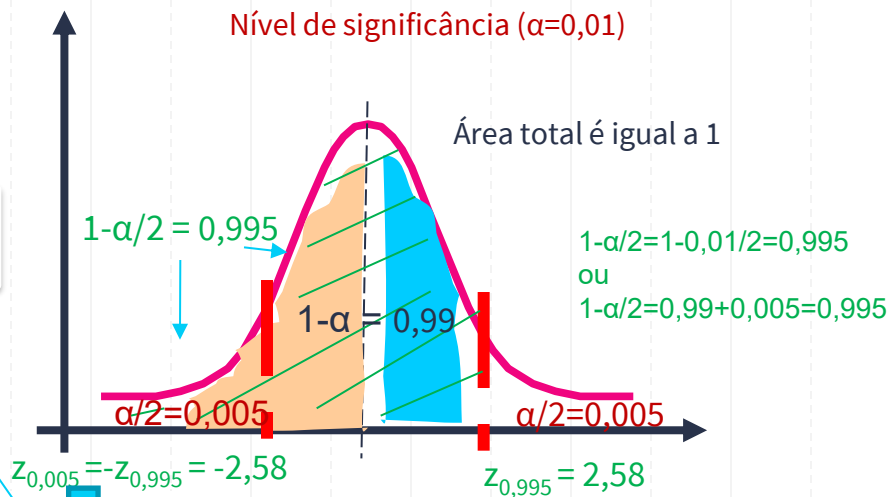
Regra: Valor-p $< \alpha \Rightarrow$ Rejeita-se H_0

Região de rejeição ou crítica:

$RR =] -\infty; -z_{0,995}] \cup [z_{0,995}; +\infty[=] -\infty; -2,5758] \cup [2,5758; +\infty[$

Cálculo do Quantil da Distribuição Normal Padrão de Probabilidade α

O nível de significância é igual a $\alpha = 0,01$, então tem-se $z_{0,995} = 2,58$ (ver Tabela)



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986

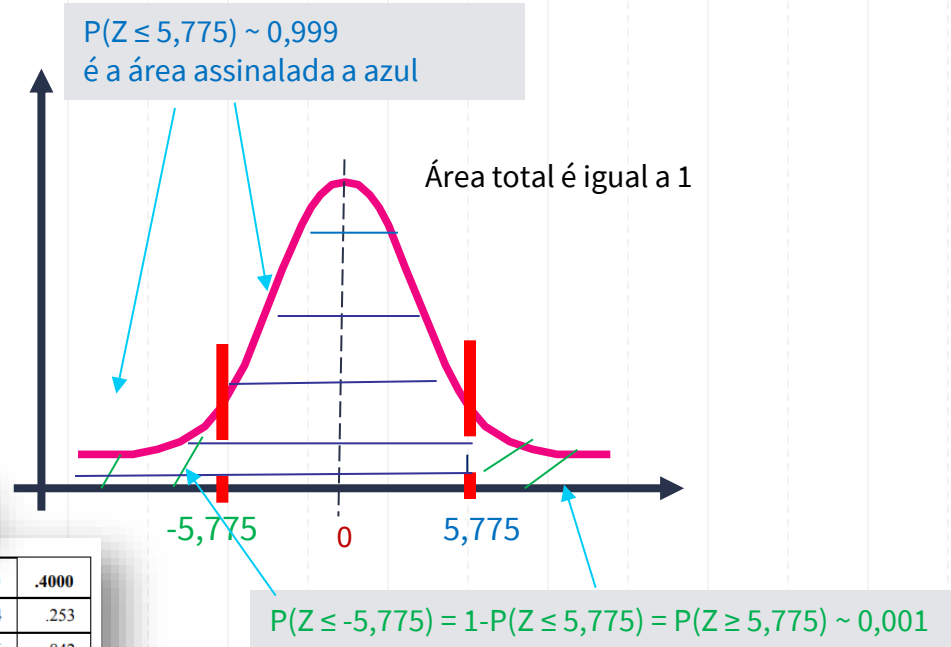
Teste bilateral: valor-p = $P(Z \leq -z_0 \text{ ou } Z \geq z_0) = P(Z \leq -z_0) + P(Z \geq z_0) = 2 \times P(Z \geq |z_0|)$

Cálculo do Valor-p quando a Estatística de Teste tem Distribuição Normal Padrão

valor-p = $P(Z \leq -5,775 \text{ ou } Z \geq 5,775)$
 $= 2 \times P(Z \geq 5,775) \sim 2 \times P(Z \geq 3,290) = 2 \times 0,0005 = 0,001$

$z_\varepsilon : P(Z > z_\varepsilon) = \varepsilon ; \quad z_{\varepsilon/2} : P(|Z| > z_{\varepsilon/2}) = \varepsilon .$

ε	.0005	.0010	.0050	.0100	.0200	.0250	.0500	.1000	.2000	.3000	.4000
z_ε	3.290	3.090	2.576	2.326	2.054	1.960	1.645	1.282	.842	.524	.253
$z_{\varepsilon/2}$	2.481	3.290	2.807	2.576	2.326	2.241	1.960	1.645	1.282	1.036	.842



45. A empresa “Trapos & Companhia” garante que pelo menos 95% das peças de fazenda que produz não têm defeito. De um grande fornecimento recebido por um armazém, foram seleccionadas 100 peças de fazenda, tendo-se apurado que sete delas apresentavam defeitos.

- a) Pode pôr-se em causa a garantia da empresa? Justifique com base no teste de hipóteses adequado.
- b) Qual a potência do teste efectuado na alínea anterior se a verdadeira proporção de peças defeituosas for de 0.1?



Exercício 45

A empresa "Trapos & Companhia" garante que pelo menos 95% das peças de fazenda que produz não têm defeito. De um grande fornecimento recebido por um armazém, foram seleccionadas 100 peças de fazenda, tendo-se apurado que sete delas apresentavam defeitos.

População de Bernoulli:

$$X \equiv \begin{cases} 1 & \text{(se a peça não é defeituosa)} \\ 0 & \text{(caso contrário)} \end{cases}$$

$$\sim \text{Ben}(\theta)$$

Proporção de peças não defeituosas

Amostra: $n = 100$ peças

$$\sum_{i=1}^{100} x_i = 100 - 7 = 93$$
$$\bar{x} = \frac{93}{100} = 0.93$$

Exercício 45 a)

- a) Pode pôr-se em causa a garantia da empresa? Justifique com base no teste de hipóteses adequado.

$$H_0: \theta \geq 0.95 \quad \text{vs} \quad H_1: \theta < 0.95 \quad (\alpha = 0.05)$$

$(\theta = 0.95)$

$$\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \stackrel{a}{\sim} N(0,1)$$

$$\text{Estatística do teste: } Z_0 = \frac{\bar{X} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \stackrel{a}{\sim} N(0,1)$$

$$W_Z = \{ z_{obs} : z_{obs} < -z_{0.05} = -1.645 \}$$

$$z_{obs} = \frac{\bar{x} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} = \frac{0.93 - 0.95}{\sqrt{\frac{0.95(0.05)}{100}}} \approx -0.9177 \notin W_Z$$

↓
Não se rejeita H_0 .

A evidência estatística corrobora a garantia da empresa.

Exercício 45 b)

$$\hookrightarrow 1 - \theta = 0.1 \Leftrightarrow \theta = 0.9$$

$\hookrightarrow$ Proporção de peças defeituosas

$$1 - \beta = P(\text{rejeitar } H_0 \mid H_0 \text{ é falsa}) = P(Z_0 < -1.645 \mid \theta = 0.9)$$

$$= P\left(\frac{\bar{X} - 0.95}{\sqrt{\frac{0.95(0.05)}{100}}} < -1.645 \mid \theta = 0.9\right) =$$

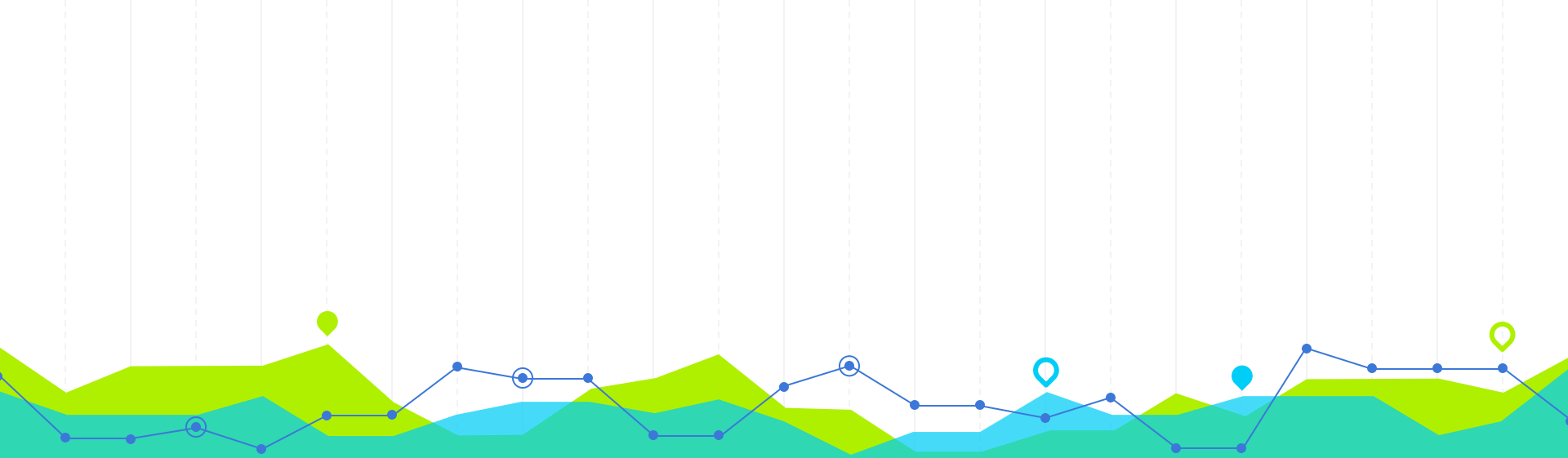
$$= P\left(\bar{X} < 0.95 - 1.645 \sqrt{\frac{0.95(0.05)}{100}} \mid \theta = 0.9\right)$$

$$= P\left(\underbrace{\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}}_{\hookrightarrow \sim N(0,1)} < \frac{0.95 - 1.645 \sqrt{\frac{0.95(0.05)}{100}} - 0.9}{\sqrt{\frac{0.9 \times 0.1}{100}}}\right) =$$

$$\approx \Phi(0.47) \underset{\substack{\downarrow \\ \text{Tabela}}}{=} 0.6808$$

A diferença para as soluções deve-se apenas ao arredondamento feitos antes de ir à tabela:

```
> pnorm((0.95-0.9-1.645*sqrt(0.95*0.05/100)) / sqrt(0.9*0.1/100))
[1] 0.6813945
```



Testes de Hipóteses para $p_1 - p_2$

Hipóteses, Estatística de Teste e Decisão

3

Teste de Hipóteses para $p_1 - p_2$: Formulário

População de Bernoulli

Proporção	$\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} N(0,1)$
Diferença de proporções	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{m} + \frac{\theta_2(1-\theta_2)}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{m} + \frac{\bar{X}_2(1-\bar{X}_2)}{n}}} \stackrel{a}{\sim} N(0,1)$
Igualdade de proporções	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \hat{\theta}(1-\hat{\theta})}} \stackrel{a}{\sim} N(0,1)$ onde $\hat{\theta} = \frac{m\bar{X}_1 + n\bar{X}_2}{m+n}$	

Teste de Hipóteses

Intervalo de Confiança

$$Z = \frac{(\bar{P}_1 - \bar{P}_2) - (p_1 - p_2)_0}{\sqrt{\bar{P}^* (1 - \bar{P}^*) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \stackrel{a}{\sim} N(0; 1), \text{ onde } \bar{P}^* = \frac{n_1 \bar{P}_1 + n_2 \bar{P}_2}{n_1 + n_2}.$$

Portanto, quando as amostras são grandes, o I. C. para $p_1 - p_2$ com $100(1 - \alpha)\%$ de confiança é dado por:

$$\left[\bar{P}_1 - \bar{P}_2 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{P}_1(1-\bar{P}_1)}{n_1} + \frac{\bar{P}_2(1-\bar{P}_2)}{n_2}}; \bar{P}_1 - \bar{P}_2 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{P}_1(1-\bar{P}_1)}{n_1} + \frac{\bar{P}_2(1-\bar{P}_2)}{n_2}} \right].$$

Teste de Hipóteses para $p_1 - p_2$

Considere duas populações Bernoulli, com parâmetros p_1 e p_2 das quais se extraíram aleatoriamente duas amostras independentes de grande dimensão n_1 e n_2 , respectivamente. Pretende-se realizar um teste de hipóteses para comparar as duas proporções amostrais p_1 e p_2 , i. e., para a diferença de proporções $p_1 - p_2$. Para este teste sabe-se que

$$Z = \frac{(\bar{P}_1 - \bar{P}_2) - (p_1 - p_2)_0}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \underset{\sim}{\sim} N(0; 1),$$

onde $(p_1 - p_2)_0$ representa o valor que se assume para $p_1 - p_2$ em H_0 .

A expressão apresentada no denominador não é conhecida, apenas se conhecendo o valor da diferença $(p_1 - p_2)$ sob H_0 . Como habitualmente este teste é realizado considerando $p_1 - p_2 = 0$, ou seja, $p_1 = p_2 = p$, e como esta proporção é desconhecida é substituída pelo seu estimador consistente. Desta forma, a estatística de teste a utilizar é:

$$Z = \frac{(\bar{P}_1 - \bar{P}_2) - (p_1 - p_2)_0}{\sqrt{\bar{P}^* \left(1 - \bar{P}^*\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \underset{\sim}{\sim} N(0; 1), \text{ onde } \bar{P}^* = \frac{n_1 \bar{P}_1 + n_2 \bar{P}_2}{n_1 + n_2}.$$

Realizou-se um estudo em duas cidades, A e B, sobre a percentagem de homens que viam o telejornal todos os dias. Na cidade A inquiriram-se aleatoriamente 150 homens tendo 54 afirmado que viam o telejornal todos os dias ao passo que na cidade B dos 200 inquiridos 80 fizeram tal afirmação.

- a) Ao nível de significância de 5%, será de admitir que a proporção de homens que vê o telejornal todos os dias é:
 - i. Diferente nas duas cidades?
 - ii. Inferior na cidade B?
 - iii. Superior na cidade B?
- b) Para cada um dos testes anteriores calcule o nível de significância a partir do qual rejeita a hipótese nula.

Probabilidades Estatística 2019 (uevora.pt)



Exercício (a) (i): Teste de Hipóteses para $p_1 - p_2$

Sejam:

- X_{1i} a v. a. que designa se o i -ésimo homem, da cidade A, afirmou ver o telejornal, $i = 1, \dots, n_1$,
- X_{2i} a v. a. que designa se o i -ésimo homem, da cidade B, afirmou ver o telejornal, $i = 1, \dots, n_2$,
- $\bar{P}_1$ a v. a. que representa a proporção de homens, da cidade A, que afirmaram ver o telejornal, em n_1 homens,
- $\bar{P}_2$ a v. a. que representa a proporção de homens, da cidade B, que afirmaram ver o telejornal, em n_2 homens.

$$n_1 = 150; \bar{p}_1 = \frac{54}{150} = 0,36; n_2 = 200 \text{ e } \bar{p}_2 = \frac{80}{200} = 0,4.$$

a) $\alpha = 5\%$.

Exercício (a) (i): Teste de Hipóteses para $p_1 - p_2$

i) $p_1 \neq p_2$?

Hipóteses:

$$H_0: p_1 = p_2 \text{ vs. } H_1: p_1 \neq p_2$$

$$\Leftrightarrow H_0: p_1 - p_2 = 0 \text{ vs } H_1: p_1 - p_2 \neq 0 \text{ (teste bilateral).}$$

Estatística de teste:

Estatística de Teste:

$$Z = \frac{(\bar{P}_1 - \bar{P}_2) - (p_1 - p_2)_0}{\sqrt{\bar{P}^* (1 - \bar{P}^*) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0; 1), \text{ onde } \bar{P}^* = \frac{n_1 \bar{P}_1 + n_2 \bar{P}_2}{n_1 + n_2}.$$

$$\bar{p}^* = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2} = \frac{150 \times 0,36 + 200 \times 0,4}{150 + 200} = 0,3829,$$

$$z_{obs} = \frac{(0,36 - 0,4) - 0}{\sqrt{0,3829(1 - 0,3829) \left(\frac{1}{150} + \frac{1}{200} \right)}} = -0,7619.$$

[ProbabilidadesEstatistica 2019 \(uevora.pt\)](#)

Pela tabela $z_{1-\frac{\alpha}{2}} = z_{0,975} = 1,96$.

Logo, $R.A.:]-1,96; 1,96[$ e $R.R.:]-\infty; -1,96] \cup [1,96; +\infty[$.

Como $z_{obs} \in R.A.$ não rejeitar H_0 . Portanto, ao nível de significância de 5%, não existe evidência estatística de que a proporção de homens que afirmam ver o telejornal todos os dias é significativamente diferente nas duas cidades.

Exercício (a) (ii): Teste de Hipóteses para $p_1 - p_2$

ii) $p_1 > p_2$?

Hipóteses:

$H_0: p_1 \leq p_2$ vs. $H_1: p_1 > p_2$

$\Leftrightarrow H_0: p_1 - p_2 \leq 0$ vs $H_1: p_1 - p_2 > 0$ (teste unilateral direito).

Estatística de teste:

Estatística de Teste:

$$Z = \frac{(\bar{P}_1 - \bar{P}_2) - (p_1 - p_2)_0}{\sqrt{\bar{P}^* \left(1 - \bar{P}^*\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \underset{\sim}{\sim} N(0; 1), \text{ onde } \bar{P}^* = \frac{n_1 \bar{P}_1 + n_2 \bar{P}_2}{n_1 + n_2}.$$

Como a estatística de teste é a mesma da alínea anterior, $z_{obs} = -0,7619$.

Pela tabela $z_{1-\alpha} = z_{0,95} = 1,645$. Logo, $R.A.:]-\infty; 1,645[$ e $R.R.: [1,645; +\infty[$.

Como $z_{obs} \in R.A.$ não rejeitar H_0 . Portanto, ao nível de significância de 5%, não existe evidência estatística de que a proporção de homens que afirmam ver o telejornal todos os dias seja significativamente inferior na cidade B.

Exercício (a) (iii): Teste de Hipóteses para $p_1 - p_2$

iii) $p_1 < p_2$?

Hipóteses:

$H_0: p_1 \geq p_2$ vs. $H_1: p_1 < p_2$

$\Leftrightarrow H_0: p_1 - p_2 \geq 0$ vs $H_1: p_1 - p_2 < 0$ (teste unilateral esquerdo).

Estatística de teste:

Estatística de Teste:

$$Z = \frac{(\bar{P}_1 - \bar{P}_2) - (p_1 - p_2)_0}{\sqrt{\bar{P}^* (1 - \bar{P}^*) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \underset{\sim}{\sim} N(0; 1), \text{ onde } \bar{P}^* = \frac{n_1 \bar{P}_1 + n_2 \bar{P}_2}{n_1 + n_2}.$$

Como a estatística de teste é a mesma da alínea anterior, $z_{obs} = -0,7619$.

Pela tabela $z_{1-\alpha} = z_{0,95} = 1,645$. Logo, $R. A. :]-1,645; +\infty[$ e $R. R. :]-\infty; -1,645]$.

Como $z_{obs} \in R. A.$ não rejeitar H_0 . Portanto, ao nível de significância de 5%, não existe evidência estatística de que a proporção de homens que afirmam ver o telejornal todos os dias seja significativamente superior na cidade B.

Exercício (b): Teste de Hipóteses para $p_1 - p_2$

b) i) valor $p = 2 \times P(Z \geq |z_{obs}|) = 2 \times P(Z \geq 0,7619) = 2 \times (1 - \Phi(0,7619))$
 $= 2 \times (1 - 0,7769) = 0,4461$.

Decisão (pela valor-p):

A hipótese $H_0: p_1 - p_2 = 0$ é rejeitada para níveis de significância superiores ou iguais a 44,61%, logo não existe evidência de que as proporções sejam diferentes, para qualquer nível de significância usual (1%, 5%, 10%).

ii) valor $p = P(Z \geq z_{obs}) = P(Z \geq -0,7619) = 1 - \Phi(-0,7619) = \Phi(0,7619) = 0,7769$.

Alternativa, com base no *valor-p* bilateral calculado na alínea i): substituindo em H_1 p_1 e p_2 por $\bar{p}_1$ e $\bar{p}_2$, respectivamente, $\bar{p}_1 - \bar{p}_2 = 0,36 - 0,4 < 0$ dá uma proposição falsa. Logo,

$$\text{valor } p_{uni} = 1 - \frac{0,4461}{2} = 0,7769.$$

A hipótese $H_0: p_1 - p_2 \leq 0$ é rejeitada para níveis de significância superiores ou iguais a 77,69%. logo não existe evidência de que a proporção na cidade A seja superior à verificada na cidade B.

iii) valor $p = P(Z \leq z_{obs}) = P(Z \leq -0,7619) = \Phi(-0,7619) = 1 - \Phi(0,7619)$
 $= 1 - 0,7769 = 0,2231$.

Alternativa, com base no *valor-p* bilateral calculado na alínea i): substituindo em H_1 p_1 e p_2 por $\bar{p}_1$ e $\bar{p}_2$, respectivamente, $\bar{p}_1 - \bar{p}_2 = 0,36 - 0,4 < 0$ dá uma proposição verdadeira. Logo,

$$\text{valor } p_{uni} = \frac{0,4461}{2} = 0,2231.$$

A hipótese $H_0: p_1 - p_2 \geq 0$ é rejeitada para níveis de significância superiores ou iguais a 22,31%, logo não existe evidência de que a proporção na cidade A seja inferior à verificada na cidade B.

48. O Ministério das Finanças está interessado em averiguar se a proporção de declarações de IRS enviadas pela *Internet* com erros de preenchimento é menor do que aquela que se verifica nas declarações entregues nos balcões das repartições de Finanças. Uma amostra de 500 declarações enviadas pela *Internet* evidenciou que 125 continham erros de preenchimento, enquanto uma amostra de 450 declarações entregues nos balcões revelou 128 mal preenchidas. Efectue o teste adequado e interprete o resultado obtido.



Exercício 48

Duas populações de Bernoulli:

Declarações enviadas pela internet:

$$X = \begin{cases} 1 & \text{(se a declaração tem erros)} \\ 0 & \text{(caso contrário)} \end{cases}$$

$$X \sim \text{Ben}(\theta_X)$$

Declarações entregues nos balcões:

$$Y = \begin{cases} 1 & \text{(se a declaração tem erros)} \\ 0 & \text{(caso contrário)} \end{cases}$$

$$Y \sim \text{Ben}(\theta_Y)$$

Exercício 48

Amostras:

$$m_x = 500$$

$$\sum_{i=1}^{500} x_i = 125$$

$$\bar{x} = \frac{125}{500} = 0.25$$

$$m_y = 450$$

$$\sum_{i=1}^{450} y_i = 128$$

$$\bar{y} = \frac{128}{450} = 0.28(4)$$

$$H_0: \theta_x - \theta_y \geq 0 \quad \text{vs} \quad H_1: \theta_x - \theta_y < 0$$

$(\theta_x - \theta_y = 0)$

Duas populações de Bernoulli, igualdade de proporções

$$\frac{\bar{X} - \bar{Y} - (\theta_x - \theta_y)}{\sqrt{\left(\frac{1}{m_x} + \frac{1}{m_y}\right) \hat{\theta}(1-\hat{\theta})}} \sim N(0,1)$$

$$\hat{\theta} = \frac{m_x \bar{X} + m_y \bar{Y}}{m_x + m_y}$$

Exercício 48

$$\text{Sob } H_0, Z_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{1}{n_X} + \frac{1}{n_Y}\right) \hat{\theta}(1-\hat{\theta})}} \stackrel{a}{\sim} N(0,1)$$

$$W_2 = \} \mathcal{Z}_{obs}: \mathcal{Z}_{obs} < - \mathcal{Z}_{0.05} = -1.645 \{$$

$$\hat{\theta} = \frac{n_X \bar{x} + n_Y \bar{y}}{n_X + n_Y} = \frac{125 + 128}{500 + 450} \simeq 0.2663$$

$$\mathcal{Z}_{obs} = \frac{(0.25 - 0.284) - 0}{\sqrt{\left(\frac{1}{500} + \frac{1}{450}\right) 0.2663(1-0.2663)}} = -1.18376 \notin W_2$$

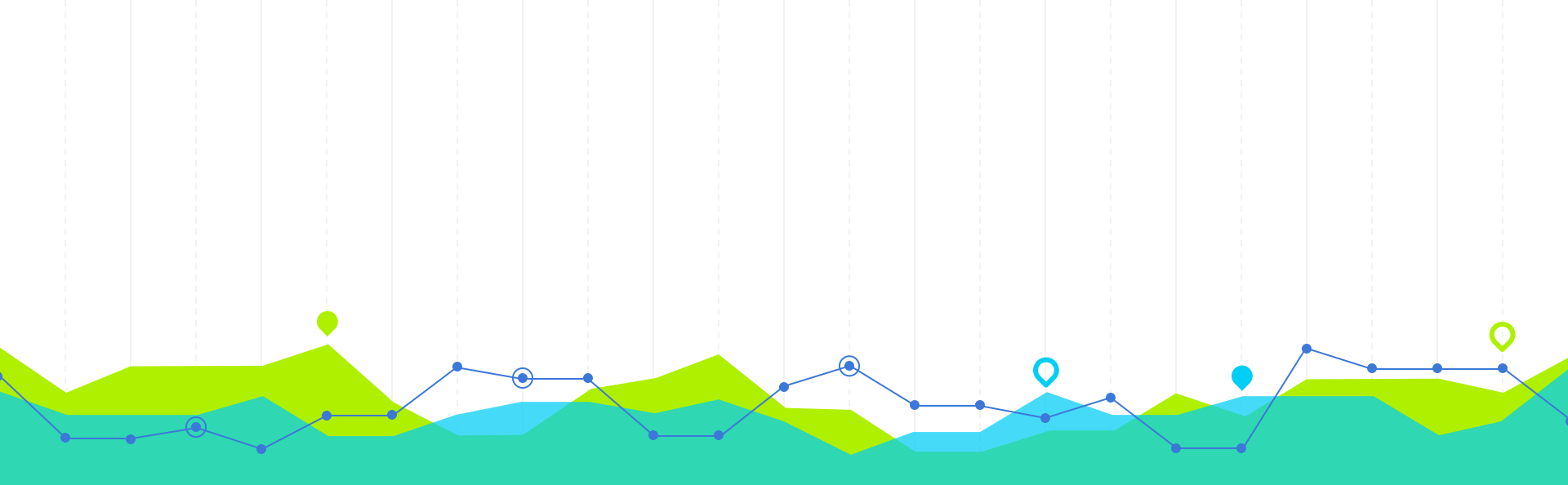
Não se rejeita H_0 .

Exercício 48

De acordo com o teste efetuado, a evidência estatística parece não corroborar a hipótese de que as declarações enviadas pela internet contêm menos erros.

Nota: Para obter o resultado das soluções basta arredondar as contas apenas no ultimo passo.

```
> theta <- (125 + 128) / (500 + 450)
> (0.25 - 128 / 450) / sqrt((1 / 500 + 1 / 450)* theta*(1 - theta))
[1] -1.199211
```



Teste de Hipóteses de Ajustamento do Q-Q

Hipóteses, Estatística de Teste e Decisão

4

4. O tabagismo é um fator de risco para neoplasia gástrica. Pretende-se saber se se pode considerar que a ocorrência de neoplasia gástrica, em fumadores, é igualmente provável na região pilórica, no corpo gástrico e na região do cárdia. Observada uma amostra aleatória constituída por 161 indivíduos com neoplasia gástrica e fumadores de 20 cigarros/dia pelo menos durante 20 anos, obteve-se a seguinte tabela de frequências:

Região pilórica	Corpo gástrico	Região do cárdia
45	54	62

Realizado o teste estatístico adequado, obteve-se o *output*:

Test Statistics	
	Neoplasia gástrica
Chi-Square ^a	2,696
df	2
Asymp. Sig.	,260
a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is	

- 4.1. Identifique, justificando, o teste estatístico utilizado.
- 4.2. Formule as hipóteses estatísticas associadas ao teste.
- 4.3. Calcule as frequências esperadas sob a hipótese nula, e complete o output.
- 4.4. Indique o valor observado da estatística de teste e a forma como foi obtido.
- 4.5. O que pode afirmar ao nível de significância de 5%?



Exercício 4: Variável

Localização_neoplasia

- Localização de neoplasia gástrica (1-Região pilórica, 2-Corpo gástrico, 3-Região do cárdia) em fumadores
- Qualitativa nominal

Freq

- Frequência absoluta

Exercícios 4.1. e 4.2: Teste de Ajustamento do Qui-Quadrado

Hipóteses

$H_0: p_1 = p_2 = p_3 = 1/3 = 33,3\%$ (“Ocorrência de neoplasia gástrica é igualmente provável nas 3 regiões”)

Versus

Dados

$n = 161$

p_i = probabilidade da ocorrência de neoplasia gástrica na i -ésima região, $i = 1, 2, 3$

H_1 : Pelo menos uma destas probabilidades regista outro valor diferente na população (ou “Ocorrência de neoplasia gástrica não é igualmente provável nas 3 regiões”)

Objetivo do Teste de Ajustamento do Qui-Quadrado:

- ✓ Pretende-se saber se a ocorrência de neoplasia gástrica, em fumadores, é igualmente provável em três regiões do corpo (região pilórica, corpo gástrico e região do cárdia). De outra forma, pretende-se testar se a frequência dessa doença é igual nas 3 regiões do corpo.

Exercícios 4.3. e 4.4: Teste de Ajustamento do Qui-Quadrado

Formulário

Estatística de teste

$$\sum_i \frac{(O_i - E_i)^2}{E_i} \underset{\text{aprox}}{\sim} \chi^2_{(k-1)}$$

Teste de Ajustamento: $Q = \sum_{j=1}^m \frac{(N_j - fe_j)^2}{fe_j} \underset{a}{\sim} \chi^2(m-1)$

Com estimação de k parâmetros para obter as estimativas $\hat{p}_{\circ j}$: $\chi^2_{(m-k-1)}$

Pela fórmula:

$$\text{VOE} = (-8,67)^2 / 53,67 + 0,33^2 / 53,67 + 8,33^2 / 53,67 = 2,695$$

Resposta: E_i mínimo é 53,67

Região	Freq. obser. (O_i)	Freq. esper. ($E_i = n \times p_i$)	Resíduos ($O_i - E_i$)
1 - Região pilórica	45	$161 \times 1/3 = 53,67$	-8,67
2 - Corpo gástrico	54	53,67	0,33
3 - Região do cárdia	62	53,67	8,33
Total	161		

Condições de Aplicabilidade dos Testes do Qui-Quadrado:

- As frequências esperadas devem ser ≥ 5 .
- No caso de tal não se verificar, então pelo menos 80% das frequências esperadas ≥ 5 e todas > 1

A Condição de Aplicabilidade dos Testes do Qui-Quadrado é satisfeita:

- Todas as frequências esperadas $E_i \geq 5$.

Exercício 4.5: Teste de Ajustamento do Qui-Quadrado

Decisão (para $\alpha = 0,05$)

Pelo valor crítico: $\text{VOE} = 2,695 < \chi^2_{0,95;2} = 5,99$

Tabela da Distribuição do Qui-Quadrado

Quantil de probabilidade $1-\alpha$ da distribuição Qui-Quadrado

Regra de decisão pelo valor crítico ou região de rejeição (RR):

$\begin{cases} \text{VOE} \geq \chi^2_{1-\alpha} \\ \text{VOE} \in \text{RR} = [\chi^2_{1-\alpha}; +\infty[\end{cases} \Rightarrow \text{Rejeita-se } H_0 \text{ para } \alpha$

Região de rejeição ou crítica:

$2,695 \notin \text{RR} = [\chi^2_{95;2}; +\infty[= [5,99; +\infty[$

Regra de decisão pelo valor-p:

$\text{Valor-p} = P(X^2 \geq \text{VOE}) < \alpha \Rightarrow \text{Rejeita-se } H_0 \text{ para } \alpha$

Pelo valor-p: $\text{valor-p} = P(X^2 \geq 2,695) = 0,260 > 0,05$

Cálculo do Quantil da Distribuição Qui-Quadrado de Probabilidade $1-\alpha$ e com $n-1$ g.l.'s

Nível de confiança ($1-\alpha=0,95$)

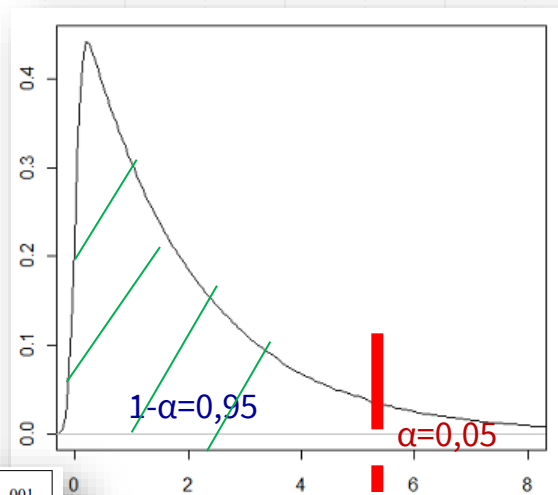
Nível de significância ($\alpha=0,05$)

Área total é igual a 1

O nível de significância é igual a $\alpha = 0,05$, então tem-se $1-\alpha = 0,95$

$\chi^2_{0,95;2} = 5,991$ (ver tabela)

$$\chi^2_{n,\varepsilon} : P(X > \chi^2_{n,\varepsilon}) = \varepsilon$$



$$\chi^2_{0,95;2} = 5,991$$

ε	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005	.001
n														
1	.000	.000	.001	.004	.016	.102	.455	1.323	2.706	3.841	5.024	6.635	7.879	10.827
2	.010	.020	.051	.103	.211	.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597	13.815
3	.072	.115	.216	.352	.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	.207	.297	.484	.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860	18.466
5	.412	.554	.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.832	15.086	16.750	20.515
6	.676	.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548	22.457
7	.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278	24.321
8	1.344	1.647	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955	26.124
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188	29.588

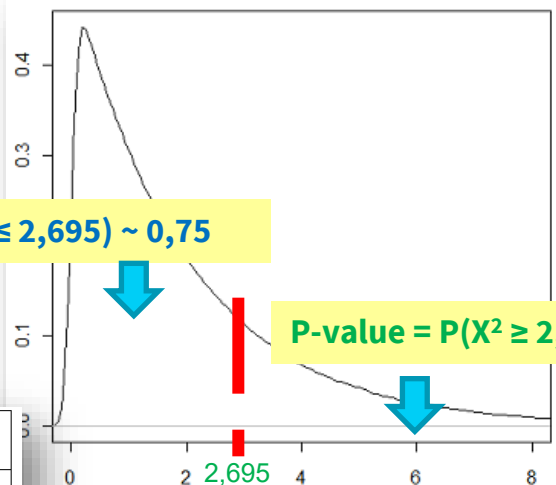
Cálculo do Valor-p quando a Estatística de Teste tem Distribuição Qui-Quadrado

$$\text{valor-p} = P(X^2 \geq 2,695) \sim P(X^2 \geq 2,773) = 0,25$$

$$\chi^2_{n,\varepsilon} : P(X > \chi^2_{n,\varepsilon}) = \varepsilon$$

ε	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005	.001
n														
1	.000	.000	.001	.004	.016	.102	.455	1.323	2.706	3.841	5.024	6.635	7.879	10.827
2	.010	.020	.051	.103	.211	.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597	13.815
3	.072	.115	.216	.352	.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	.207	.297	.484	.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860	18.466
5	.412	.554	.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.832	15.086	16.750	20.515
6	.676	.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548	22.457
7	.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278	24.321
8	1.344	1.647	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955	26.124
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188	29.588

Área total é igual a 1



$$P(X^2 \leq 2,695) \sim 0,75$$

$$\text{P-value} = P(X^2 \geq 2,695) \sim 0,25$$

Regra de decisão pelo valor-p:

Valor-p = $P(X^2 \geq \text{VOE}) < \alpha \Rightarrow$ Rejeita-se H_0 para α

1. Com o objectivo de remodelar determinado centro comercial, realizou-se uma pesquisa sobre o movimento de entradas e saídas pelas suas três portas. Inquiriu-se qual a porta de entrada para uma amostra aleatória de 201 pessoas:

Entrada	1	2	3
N.º de pessoas	83	62	56

Foi afirmado que não havia preferência por qualquer uma das três entradas. Comente para uma dimensão de 0.05.



Exercício 1

X = Porta escolhida para entrar no centro comercial

$$D_X = \{1, 2, 3\} \quad n = 201 \text{ pessoas}$$

$$P_j = P(X = j) \quad (j = 1, 2, 3)$$

$$H_0: P_1 = P_2 = P_3 = \frac{1}{3} \quad H_1: \exists j: P_j \neq \frac{1}{3} \quad (j = 1, 2, 3)$$

$$P_{0j} = P(X = j | H_0) = \frac{1}{3} \quad (j = 1, 2, 3)$$

Exercício 1

$$Q = \sum_{j=1}^m \frac{(N_j - fe_j)^2}{fe_j} \approx \chi^2_{(m-1)} = \chi^2(2)$$

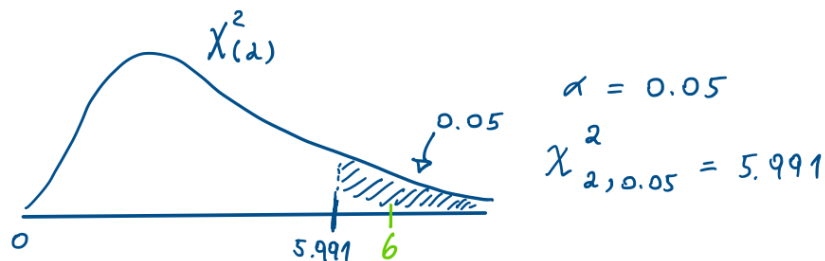
$m = m^\circ$ de categorias

Porta N_j $fe_j = m P_{0j}$

$$\left. \begin{array}{lll} 1 & 83 & 201 \times \frac{1}{3} = 67 > 5 \\ 2 & 62 & 201 \times \frac{1}{3} = 67 > 5 \\ 3 & 56 & 201 \times \frac{1}{3} = 67 > 5 \end{array} \right\} m = 3$$

$$Q_{obs} = \frac{(83 - 67)^2}{67} + \frac{(62 - 67)^2}{67} + \frac{(56 - 67)^2}{67} = 6$$

Exercício 1



$$W_{\alpha} = \{Q_{obs} : Q_{obs} > \chi^2_{2,0.05} = 5.991\}$$

$$Q_{obs} = 6 \in W_{\alpha} \text{ logo rejeita-se } H_0.$$

Conclusão : Tendo como referência um teste de dimensão 0.05 a hipótese nula é rejeitada, o que significa que a evidência estatística presente na amostra recolhida não é favorável à afirmação enunciada.

Nota : `> pchisq(6, 2, lower.tail = FALSE)`
[1] 0.04978707

$$p_{obs} = 0.0498 < \alpha = 0.05 \text{ logo rejeita-se } H_0.$$

Com $\alpha = 0.01$ já não se rejeitaria H_0 .

4. O número de erros de impressão por página de um livro é frequentemente considerado uma variável aleatória de Poisson. A contagem dos erros de impressão em 100 páginas de um livro deu o seguinte resultado:

N.º de erros	0	1	2	3
N.º de páginas	65	25	8	2

- a) Teste a hipótese, ao nível de 0.01, de que o número de erros por página é uma variável aleatória de Poisson de parâmetro $\lambda = 0.4$.



Exercício 4 a)

a) $X \equiv$ nº de erros por página
amostra: $n = 100$ páginas

$$H_0: X \sim \text{Poi}(0.4) \quad H_1: X \not\sim \text{Poi}(0.4) \quad (\alpha = 0.01)$$

erros N_j $n p_{0j}$ sob H_0 , $f_x(x) = \frac{e^{-0.4} 0.4^x}{x!} \quad (x=0,1,2,\dots)$

$$0 \quad 65 \quad 67.03$$

$$1 \quad 25 \quad 26.81$$

$$2 \quad 8 \quad 5.36$$

$$\geq 3 \quad 2 \quad 0.8$$

< 5 $\rightarrow$ Temos que agrupar estas duas categorias:
 $m = 3 \quad N_3 = 10$
 $m p_{03} = 5.36 + 0.8 = 6.16$

Exercício 4 a)

$$P_{0j} = P(X = j-1 | H_0) = \frac{e^{-0.4} 0.4^{j-1}}{(j-1)!} \quad (j = 1, 2, 3)$$

$$P_{04} = P(X \geq 3 | H_0) = 1 - P_{01} - P_{02} - P_{03}$$

Assim sendo:

$$n P_{01} = 100 P(X = 0 | H_0) = 100 \times \underbrace{0.6703}_{\text{Tabela 2}} = 67.03$$

$$n P_{02} = 100 P(X = 1 | H_0) = 100 \times \underbrace{0.2681} = 26.81$$

$$n P_{03} = 100 P(X = 2 | H_0) = 100 \times \underbrace{0.0536} = 5.36$$

$$\begin{aligned} n P_{04} &= 100 P(X \geq 3 | H_0) = 100 \times (1 - P_{01} - P_{02} - P_{03}) = \\ &= 100 (1 - 0.6703 - 0.2681 - 0.0536) = 100 \times 0.008 = 0.8 \end{aligned}$$

Exercício 4 a)

A Hipótese que realmente se testa é a hipótese "afarentada":

$$H'_0: \begin{cases} P(X = x) = \frac{e^{-0.4} 0.4^x}{x!} \quad (x = 0, 1) \\ P(X \geq 2) = 1 - P(X=0) - P(X=1) \end{cases}$$

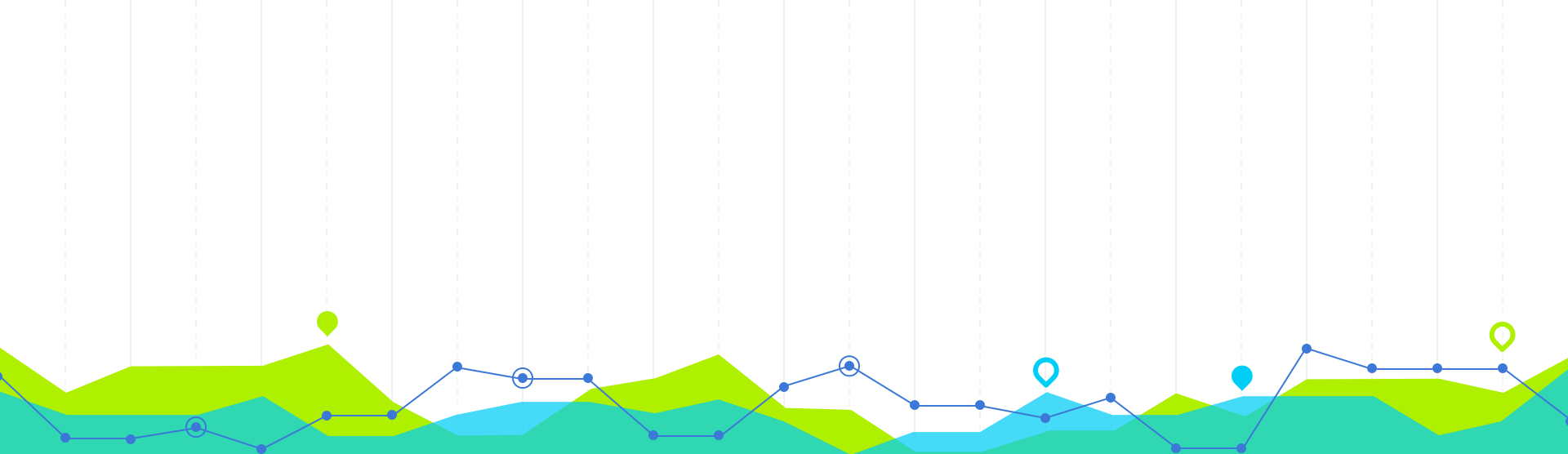
$$Q = \sum_{j=1}^3 \frac{(N_j - m p_{0j})^2}{m p_{0j}} \approx \chi^2_{(m-1)} = \chi^2_{(2)}$$

$$\chi^2_{2, 0.01} = 9.21$$

$$W_Q = \{Q_{\text{obs}} : Q_{\text{obs}} > 9.21\}$$

Exercício 4

$$\begin{aligned} Q_{obs} &= \frac{(65 - 67.03)^2}{67.03} + \frac{(25 - 26.81)^2}{26.81} + \frac{(10 - 6.16)^2}{6.16} = \\ &= 2.58 \notin W_\alpha \text{ logo não se rejeita } H_0'. \end{aligned}$$



Teste de Hipóteses de Independência do Q-Q

Hipóteses, Estatística de Teste e Decisão

5

5. Um inspetor de qualidade recolheu uma amostra de 176 produtos alimentares num centro de distribuição. Sabendo que cada produto pode ser proveniente de uma de três fábricas e pode ou não estar contaminado; o inspetor avaliou todos os produtos e obteve os seguintes resultados:

	Fábrica A	Fábrica B	Fábrica C	Total
Contaminado	8	15	11	34
Não contaminado	55	67	20	142
Total	63	82	31	176

Pode-se afirmar que o facto de um produto estar contaminado é independente da sua fábrica de origem, considerando $\alpha = 0,01$?

[Adaptado da fonte: <https://www.ime.unicamp.br/~veronica/Coordenadas1s/aula8pr.pdf>]



Formulário

$$\text{Teste de Independência: } Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - fe_{ij})^2}{fe_{ij}} \sim \chi^2_{((r-1)(s-1))}$$

Exercício: Teste de Independência do Qui-Quadrado

Hipóteses

H_0 : As variáveis são independentes

Versus

H_1 : As variáveis não são independentes

Estatística de teste

$$\sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \underset{aprox}{\sim} \chi^2_{(1-1)(c-1)}$$

Decisão

Pelo valor crítico: Valor da Estatística de Teste = 7,024 não pertence a $RR = [\chi^2_{0,99;2}; +\infty[= [9,21; +\infty[$

Pelo valor-p: valor-p = 0,030 > 0,01

Não se rejeita H_0 para $\alpha = 1\%$. Assim, não existe evidência estatística para afirmar que as variáveis não são independentes para $\alpha = 1\%$.

Dados

N = 176

VOE = 7,024

Valor-p = 0,03

$\alpha = 1\% \Rightarrow 1 - \alpha = 99\%$

df = g.l. = 2

Frequências esperadas

$$E_{ij} = \frac{L_i \times C_j}{N}$$

L = total linhas e C = total colunas

N = nº total de elementos

Tabela da distribuição do Qui-Quadrado

Exercício: Teste de Independência do Qui-Quadrado

Estatística de Teste:

$$\sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \underset{\text{aprox}}{\sim} \chi^2_{(l-1)(c-1)}$$

Frequências esperadas

$$E_{ij} = \frac{L_i \times C_j}{N}$$

L = total linhas e C = total colunas
N = n° total de elementos

Contaminado * Fábrica Tabulação cruzada

			Fábrica			Total
			Fábrica A	Fábrica B	Fábrica C	
Contaminado	Sim	Contagem	8	15	11	34
		Expected Count	12,2	15,8	6,0	34,0
	Não	Contagem	55	67	20	142
		Expected Count	50,8	66,2	25,0	142,0
Total		Contagem	63	82	31	176
		Expected Count	63,0	82,0	31,0	176,0

Valor da Estatística de Teste

Testes de chi-quadrado

	Valor	df	Sig. Assint. (2 lados)
Chi-quadrado de Pearson	7,024 ^a	2	,030
Razão de probabilidade	6,450	2	,040
Associação Linear por Linear	6,099	1	,014
N de Casos Válidos	176		

A Condição de Aplicabilidade dos Testes do Qui-Quadrado é satisfeita:

- Todas as frequências esperadas $E_{ij} \geq 5$.

a. 0 células (0,0%) esperam contagem menor do que 5. A contagem mínima esperada é 5,000.

Teste de Independência do Qui-Quadrado

Decisão (para $\alpha = 0,01$)

Pelo valor crítico: $VOE = 7,024 > \chi^2_{0,99;2} = 9,21$

Tabela da Distribuição do Qui-Quadrado

Quantil de probabilidade $1-\alpha$ da distribuição do Qui-Quadrado

Regra de decisão pelo valor crítico ou região de rejeição (RR):

$\{VOE \geq \chi^2_{1-\alpha}$
 $(VOE \in RR = [\chi^2_{1-\alpha}; +\infty[\Rightarrow \text{Rejeita-se } H_0 \text{ para } \alpha$

Região de rejeição ou crítica:

$7,024$ não pertence a $RR = [\chi^2_{0,99;2}; +\infty[= [9,21; +\infty[$

Pelo valor-p: valor-p = $0,03 > 0,01$

Regra de decisão pelo valor-p:

Valor-p = $P(X^2 \geq VOE) < \alpha \Rightarrow \text{Rejeita-se } H_0 \text{ para } \alpha$

Não se rejeita-se H_0 para $\alpha = 0,01$. Assim, não existe evidência estatística para afirmar que as variáveis não são independentes para $\alpha = 1\%$.

Cálculo do Quantil da Distribuição Qui-Quadrado de Probabilidade $1-\alpha$ e com $(l-1) \times (c-1)$ g.l.'s

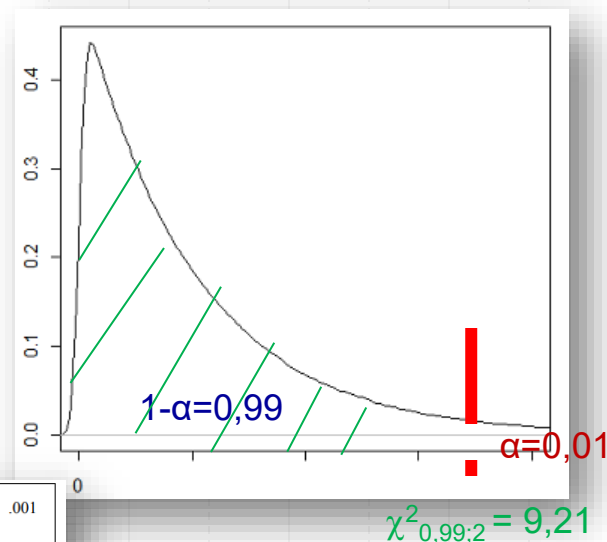
Nível de confiança ($1-\alpha=0,99$)

Nível de significância ($\alpha=0,01$)

Área total é igual a 1

O nível de significância é igual a $\alpha = 0,01$, então tem-se $1-\alpha = 0,99$

$\chi^2_{0,99;2} = 9,21$ (ver tabela)



$$\chi^2_{n,\varepsilon} : P(X > \chi^2_{n,\varepsilon}) = \varepsilon$$

ε	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005	.001
n														
1	.000	.000	.001	.004	.016	.102	.455	1.323	2.706	3.841	5.024	6.635	7.879	10.827
2	.010	.020	.051	.103	.211	.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597	13.815
3	.072	.115	.216	.352	.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	.207	.297	.484	.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860	18.466
5	.412	.554	.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.832	15.086	16.750	20.515
6	.676	.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548	22.457
7	.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278	24.321
8	1.344	1.647	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955	26.124
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188	29.588

Regra de decisão pelo valor-p:

Valor-p = $P(X^2 \geq \text{VOE}) < \alpha \Rightarrow$ Rejeita-se H_0 para α

Cálculo do Valor-p quando a Estatística de Teste tem Distribuição Qui-Quadrado

Área total é igual a 1

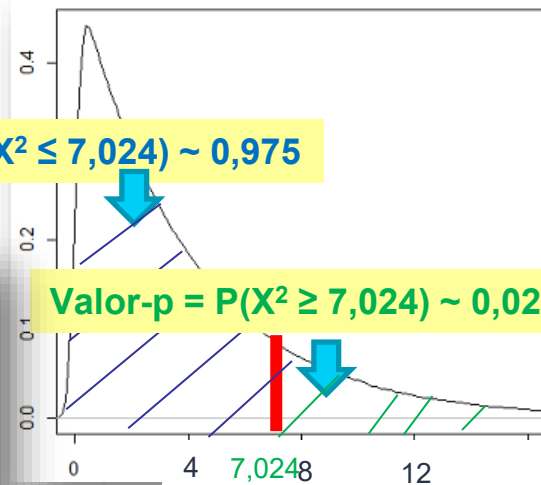
valor-p = $P(X^2 \geq 7,024) \sim P(X^2 \geq 7,378) = 0,025$

$\chi^2_{n,\epsilon} : P(X > \chi^2_{n,\epsilon}) = \epsilon$

ϵ	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005	.001
n														
1	.000	.000	.001	.004	.016	.102	.455	1.323	2.706	3.841	5.024	6.635	7.879	10.827
2	.010	.020	.051	.103	.211	.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597	13.815
3	.072	.115	.216	.352	.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	.207	.297	.484	.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860	18.466
5	.412	.554	.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.832	15.086	16.750	20.515
6	.676	.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548	22.457
7	.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278	24.321
8	1.344	1.647	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955	26.124
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188	29.588

$P(X^2 \leq 7,024) \sim 0,975$

Valor-p = $P(X^2 \geq 7,024) \sim 0,025$



Exercício: Teste de Independência do Qui-Quadrado

Condições de Aplicabilidade dos Testes do Qui-Quadrado:

- As frequências esperadas devem ser ≥ 5 .
- No caso de tal não se verificar, então pelo menos 80% das frequências esperadas ≥ 5 e todas > 1 (não é válido para tabelas 2x2).

Frequências esperadas

$$E_{ij} = \frac{L_i \times C_j}{N}$$

L = total linhas e C = total colunas
N = nº total de elementos

Verificação das condições de aplicabilidade

Neste caso, todas as células têm frequências esperadas superiores a 5.

O teste do Qui-Quadrado apenas informa sobre a independência entre variáveis, mas nada diz sobre o grau de associação existente.

Para esse efeito calculam-se **medidas de associação** tais como o coeficiente Phi, o coeficiente V de Cramer e o coeficiente de contingência.

Contaminado * Fábrica Tabulação cruzada

			Fábrica			Total
			Fábrica A	Fábrica B	Fábrica C	
Contaminado	Sim	Contagem	8	15	11	34
		Expected Count	12,2	15,8	6,0	34,0
	Não	Contagem	55	67	20	142
		Expected Count	50,8	66,2	25,0	142,0
Total		Contagem	63	82	31	176
		Expected Count	63,0	82,0	31,0	176,0

16. Uma companhia de seguros está interessada em saber se existe independência entre a realização de seguros de dois tipos: “Seguro de Vida” e “Seguro de Saúde”. Para tal analisou as apólices detidas por uma amostra casual de 200 clientes tendo-se verificado que: 90 tinham simultaneamente um seguro de vida e um seguro de saúde; 35 tinham apenas seguro de vida; 42 tinham apenas seguro de saúde; os restantes não tinham qualquer destes dois seguros. Com base num teste a 0.05, que pode concluir?



Obrigada!

Questões?

