ISEG Lisbon School of Economics & Management
Universidade de Lisboa

Estatística II

Licenciatura em Gestão
2.º Ano/1.º Semestre
2023/2024

Aulas Teóricas N.ºs 16 e 17 (Semana 9)

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 Teóricas (Semanas 1 a 5)	Aulas Teóricas (Semanas 5 a 7)	Aulas Teóricas (Semanas 7 a 9)	Aulas Teóricas (Semanas 10 a 13)
• **Capítulo 1**: Estimação	• **Capítulo 2**: Testes de Hipóteses	• **Capítulo 3**: Modelo de Regressão Linear	• **Capítulo 4**: Complementos ao Modelo de Regressão Linear

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.

6ª semana (24/10 e 26/10)

T10 - Teste de hipóteses

Hipótese simples contra hipótese composta unilateral. Testes UMP. Exemplo. Teste de hipótese simples contra hipótese composta bilateral. Exemplo.

T11 - Teste de hipóteses

Valor-p. Exemplos. Testes em universos normais: média e variância. Exemplos. Teste para 2 populações: igualdade de médias e quociente de variâncias. Exemplos.

7ª semana (31/10 a 02/11)

T12 - Teste de hipóteses

Testes em universos normais com amostras emparelhadas. Exemplo. Teste de hipóteses para grandes amostras. Aplicação ao universo de Bernoulli (média e diferença de médias). Exemplos.

T13 - Modelo de regressão linear

Introdução; modelo linear e linearizável; exemplos; Hipóteses básicas; estimação dos coeficientes da regressão pelos Mínimos Quadrados. Exemplo.

8ª semana (07/11 e 09/11)

T14 - Modelo de Regressão Linea (MRL)r

Interpretação dos parâmetros da regressão; exemplos; Resíduos MQ e regressão ajustada; Propriedades dos estimadores MQ dos coeficientes da regressão; Estimador não enviesado da variância da variável residual; Exemplo.

T15 - Modelo de regressão Linear

Coeficiente de determinação e sua interpretação. Hipótese adicional (H6) e inferência estatística sobre o modelo; Inferência sobre um parâmetro beta. Exemplos

Testes de Hipóteses para $\mu_1 - \mu_2$

Hipóteses, Estatística de Teste e Decisão
(Amostras Emparelhadas)

1

Hipóteses do Teste de Hipóteses para μ_1 - μ_2 (Amostras Emparelhadas)

Considere-se agora o caso em que as duas amostras formam um par de observações $(X_{1i};\ X_{2i})$, $i = 1, \dots, n$, ou seja, trata-se de uma amostra emparelhada. Os pares de observações são independentes e retirados de populações Normais, com médias μ_1 e μ_2 e desvios padrão σ_1 e σ_2, respectivamente. Neste caso, para testar a igualdade entre as médias populacionais, as hipóteses a formular são:

T. bilateral	**T. unilateral direito**	**T. unilateral esquerdo**
Hipóteses a testar: $H_0: \mu_1 = \mu_2\ vs\ H_1: \mu_1 \neq \mu_2$ $\Leftrightarrow H_0: \mu_1 - \mu_2 = 0\ vs$ $H_1: \mu_1 - \mu_2 \neq 0$	**Hipóteses a testar:** $H_0: \mu_1 \leq \mu_2\ vs\ H_1: \mu_1 > \mu_2$ $\Leftrightarrow H_0: \mu_1 - \mu_2 \leq 0\ vs$ $H_1: \mu_1 - \mu_2 > 0$	**Hipóteses a testar:** $H_0: \mu_1 \geq \mu_2\ vs\ H_1: \mu_1 < \mu_2$ $\Leftrightarrow H_0: \mu_1 - \mu_2 \geq 0\ vs$ $H_1: \mu_1 - \mu_2 < 0$

Para realizar o teste pretendido calcular:

- $D_i = X_{1i} - X_{2i}, \qquad i = 1, \dots, n;$
- $\overline{D} = \overline{X}_1 - \overline{X}_2 = \sum_{i=1}^{n} \frac{X_{1i}}{n} - \sum_{i=1}^{n} \frac{X_{2i}}{n} = \sum_{i=1}^{n} \frac{D_i}{n};$
- $S_D^2 = \sum_{i=1}^{n} \frac{(D_i - \bar{D})^2}{n-1}.$

Ou seja, está-se perante um teste de hipóteses para a média no caso em que a população segue uma distribuição Normal da qual se desconhece a sua variância. Esta situação já foi descrita nos capítulos anteriores (secção 8.3.2).

Desta forma, as hipóteses a testar podem escritas da seguinte forma:

Hipóteses a testar: $H_0: \mu_D = 0\ vs\ H_1: \mu_D \neq 0$	**Hipóteses a testar:** $H_0: \mu_D \leq 0\ vs\ H_1: \mu_D > 0$	**Hipóteses a testar:** $H_0: \mu_D \geq 0\ vs\ H_1: \mu_D < 0$

Um professor de estatística seleccionou aleatoriamente um grupo de 10 alunos, aprovados na disciplina pelo regime de frequências, tendo registado as suas notas nas frequências:

Aluno	1	2	3	4	5	6	7	8	9	10
1ª freq.	10	11	9	10	18	15	16	13	11	10
2ª freq.	9	12	12	12	18	14	18	12	13	10

Ao nível de significância de 5% pode afirmar que as notas médias dos alunos na 1ª frequência são superiores às obtidas na 2ª frequência? Assuma a Normalidade das notas.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício: Teste de Hipóteses para $\mu_1 - \mu_2$ (Amostras Emparelhadas)

Sejam:

- X_1 a v.a. que representa a nota do aluno na 1ª frequência,
- X_2 a v.a. que representa a nota do aluno na 2ª frequência,

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

Hipóteses:

$\alpha = 1\%, \mu_1 > \mu_2$?

$$H_0: \mu_1 \leq \mu_2 \; vs. \, H_1: \mu_1 > \mu_2$$

$$\Leftrightarrow H_0: \mu_1 - \mu_2 \leq 0 \; vs. \, H_1: \mu_1 - \mu_2 > 0$$

como as amostras são emparelhadas

$$\Leftrightarrow H_0: \mu_D \leq 0 \; vs. \, H_1: \mu_D > 0 \text{ (teste unilateral direito).}$$

Estatística de teste:

Estatística de Teste:

$$T = \frac{\overline{D} - \mu_0}{\frac{S_D}{\sqrt{n}}} \sim t_{n-1}.$$

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício: Teste de Hipóteses para μ_1 - μ_2 (Amostras Emparelhadas)

Aluno (i)	1	2	3	4	5	6	7	8	9	10
1ª freq. (x_{1i})	10	11	9	10	18	15	16	13	11	10
2ª freq. (x_{2i})	9	12	12	12	18	14	18	12	13	10
$d_i = x_{1i} - x_{2i}$	1	-1	-3	-2	0	1	-2	1	-2	0

$n = 10$; $\bar{d} = -0{,}7$ e $s_d = 1{,}4944$.

$$t_{obs} = \frac{-0{,}7 - 0}{\frac{1{,}4944}{\sqrt{10}}} = -1{,}481.$$

Decisão (pela região de rejeição):

Pela tabela $t_{n-1;\,1-\alpha} = t_{9;\,0{,}95} = 1{,}833$. Logo, $R.A.:]-\infty;\ 1{,}833[$ e $R.R.: [1{,}833; +\infty[$.

Como $t_{obs} \in R.A.$ não rejeitar H_0. Portanto, ao nível de significância de 5%, não existe evidência estatística de que as notas médias obtidas pelos alunos na 1ª frequência sejam superiores às da 2ª frequência.

ProbabilidadesEstatistica_2019 (uevora.pt)

Testes de Hipóteses para p

Hipóteses, Estatística de Teste e Decisão

2

Teste de Hipóteses para p: Formulário

População de Bernoulli

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

Estimador do Parâmetro p

Teste de Hipóteses

$$\frac{P^*-p_0}{\sqrt{\dfrac{p_0q_0}{n}}} \overset{apr}{\frown} N(0,1)$$

Intervalo de Confiança

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

<u>Exercício</u>:

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

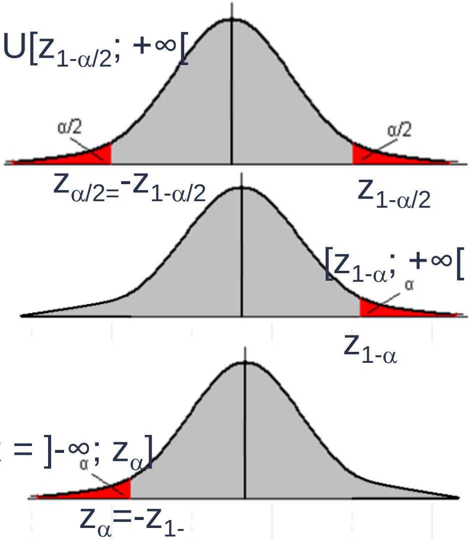
Realize um Teste de Hipóteses para testar se a percentagem de fumadores portugueses é diferente de 40% para o nível de significância $\alpha = 1\%$.

Tipos de Testes de Hipóteses para uma Proporção Populacional

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

Teste

- Bilateral: $H_0: p = p_0$, $H_1: p \neq p_0$ — $RR =]-\infty; -z_{1-\alpha/2}] \cup [z_{1-\alpha/2}; +\infty[$
- Unilateral
  - à direita: $H_0: p \leq p_0$, $H_1: p > p_0$ — $[z_{1-\alpha}; +\infty[$
  - à esquerda: $H_0: p \geq p_0$, $H_1: p < p_0$ — $RR =]-\infty; z_\alpha]$



onde p_0 é o valor numérico específico considerado em H_0 e H_1.

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

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

Hipóteses

Teste bilateral

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

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

Estatística de teste

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

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

Região de rejeição ou crítica:

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

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

Tabela da Distribuição Normal

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%.

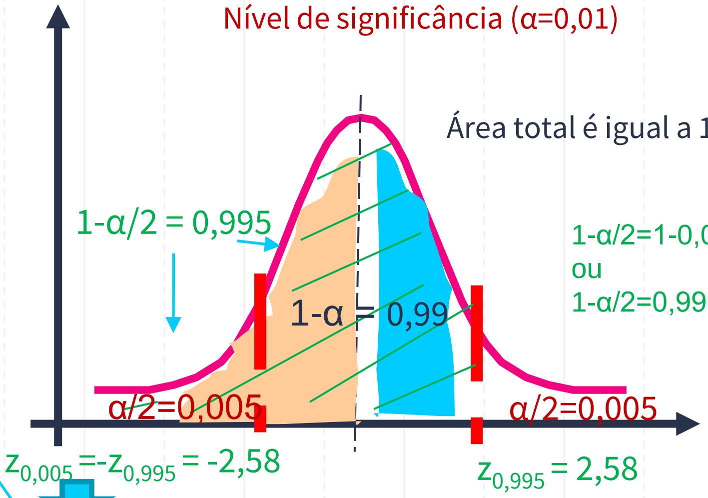
Regra: Valor-p $< \alpha \Rightarrow$ Rejeita-se H_0

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)



$1-\alpha/2=1-0,01/2=0,995$

ou

$1-\alpha/2=0,99+0,005=0,995$

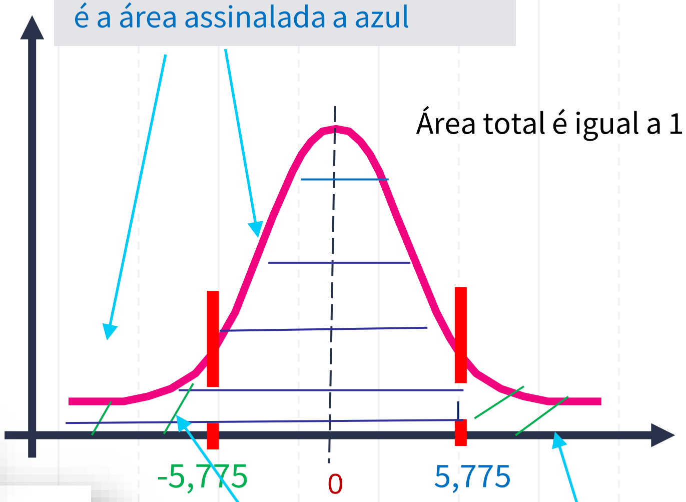
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9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

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

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

$P(Z \le 5{,}775) \sim 0{,}999$
é a área assinalada a azul

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



$P(Z \le -5{,}775) = 1 - P(Z \le 5{,}775) = P(Z \ge 5{,}775) \sim 0{,}001$

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

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
$z_{\varepsilon/2}$	3.481	3.290	2.807	2.576	2.326	2.241	1.960	1.645	1.282	1.036	.842

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

Teste de Hipóteses

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

Intervalo de Confiança

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

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

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:

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

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.

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

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

a) $\alpha = 5\%$.

ProbabilidadesEstatistica_2019 (uevora.pt)

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

Hipóteses:

Estatística de Teste:

i) $p_1 \neq p_2$?

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

Estatística de teste:

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

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

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

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

Hipóteses:

ii) $p_1 > p_2$?

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

Estatística de Teste:

Estatística de teste:

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

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.

ProbabilidadesEstatistica_2019 (uevora.pt)

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:

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

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.

ProbabilidadesEstatistica_2019 (uevora.pt)

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

Decisão (pela valor-p):

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

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 se 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 $valor\text{-}p$ bilateral calculado na alínea i): substituindo em H_1 p_1 e p_2 por $\overline{p}_1$ e $\overline{p}_2$, respectivamente, $\overline{p}_1 - \overline{p}_2 = 0{,}36 - 0{,}4 < 0$ 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 $valor\text{-}p$ bilateral calculado na alínea i): substituindo em H_1 p_1 e p_2 por $\overline{p}_1$ e $\overline{p}_2$, respectivamente, $\overline{p}_1 - \overline{p}_2 = 0{,}36 - 0{,}4 < 0$ 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.

ProbabilidadesEstatistica_2019 (uevora.pt)

4

Testes de Hipóteses para σ^2

Hipóteses, Estatística de Teste e Decisão

Teste de Hipóteses para σ^2: Formulário

Variância corrigida

$$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- **POPULAÇÕES NORMAIS**

Média	$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\dfrac{\bar{X}-\mu}{S'/\sqrt{n}} \sim t(n-1)$
Diferença de médias	$\dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} \sim N(0,1)$ $T = \dfrac{\dfrac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}}{\sqrt{\dfrac{(m-1)S_1'^2+(n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	$Z = \dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1'^2}{m}+\dfrac{S_2'^2}{n}}} \overset{a}{\sim} t(\nu)$ onde ν é o maior inteiro contido em r, $r = \dfrac{\left(\dfrac{s_1'^2}{m}+\dfrac{s_2'^2}{n}\right)^2}{\dfrac{1}{m-1}\left(\dfrac{s_1'^2}{m}\right)^2+\dfrac{1}{n-1}\left(\dfrac{s_2'^2}{n}\right)^2}$
Variância	$\dfrac{nS^2}{\sigma^2} = \dfrac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	
Relação de variâncias	$\dfrac{S_1'^2}{S_2'^2}\dfrac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$	

Exercício:
Considere-se uma amostra de dimensão $n = 9$ e variância $s^2 = 100$. Realize um Teste de Hipóteses para testar se o desvio padrão populacional é superior a 4 para o nível de significância $\alpha = 1\%$.

Exercício: Teste de Hipóteses para σ^2

Hipóteses:

Passo 0 + Passo 1

$X \sim N(\mu, \sigma^2)$ $\quad n = 9 \quad s^2 = 100$

$H_0: \sigma = 4$

$H_1: \sigma > 4$

$\alpha = 1\%$

$E(X) = \mu$

$Var(X) = \sigma^2$

Slides Professora Claúdia Nunes

Exercício: Teste de Hipóteses para σ^2

Estatística de Teste:

(2) p/ σ^2

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Passo 2

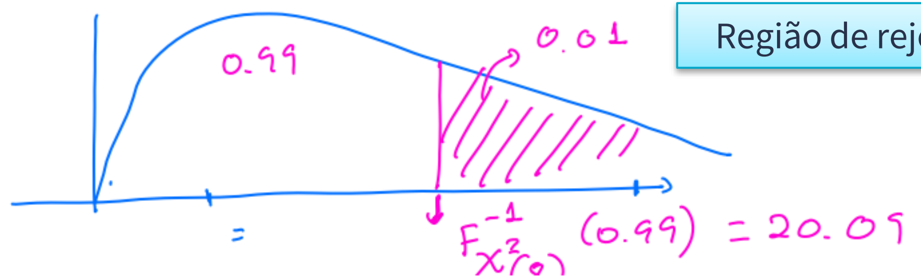
v. aleat. $T = \frac{8\, S^2}{\sigma^2} \sim \chi^2_{(8)}$ $\qquad \cancel{\frac{6 \times 25}{\sigma^2} \sim \chi^2_{(6)}}$

est. Teste $T_0 = \frac{8\, S^2}{16}$

valor observado $t_0 = \frac{8 \times 100}{16} = 50$

Exercício: Teste de Hipóteses para σ^2

Decisão (pela região de rejeição):



Região de rejeição RR = [20,09; +∞[

Como $t_0 = 50 > 20.09$, então H0 deve ser rejeitada para $\alpha = 1\%$.

para $\alpha \geq 1\%$

Decisão:

Rejeita-se H0 para α = 1%.

Existe evidência estatística para afirmar que o desvio padrão populacional é superior a 4 para α = 1%.

Cálculo dos Quantis da Distribuição Qui-Quadrado de Probabilidade 1-α e com n-1 g.l.´s

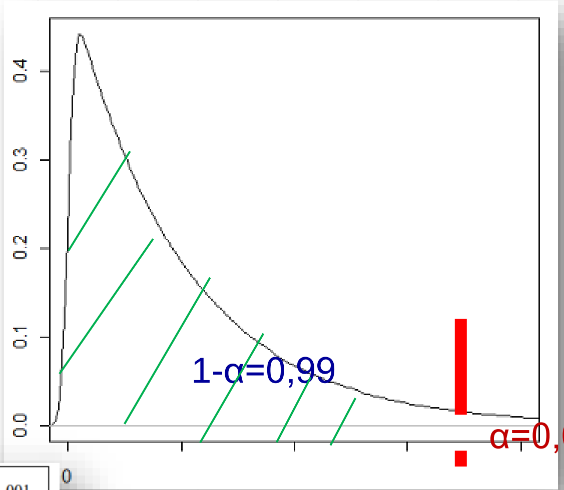
Nível de confiança (1-α=0,99)

Nível de significância (α=0,01)

Área total é igual a 1

O nível de significância é α = 0,01

Logo, pretende-se calcular o quantil da distribuição Qui-Quadrado de probabilidade 1-α = 0,99
$\chi^2_{0,99;8}$ = 20,09 (ver tabela)



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

n \ ε	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005	.001
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 = $P(X^2 \geq VOE) < \alpha \Rightarrow$ Rejeita-se H_0 para α

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

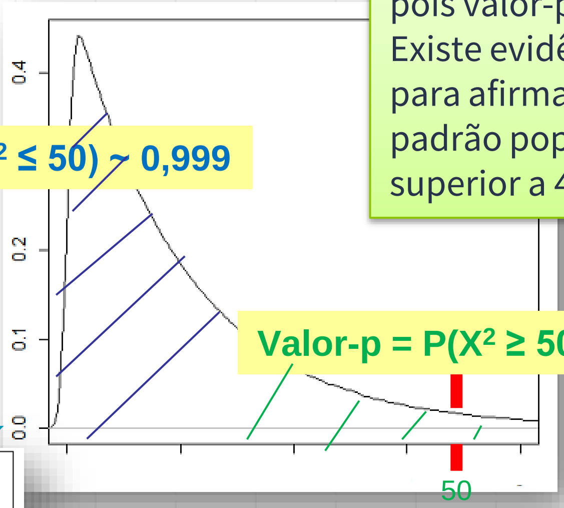
Decisão (pelo valor-p):

valor-p = $P(X^2 \geq 50) \sim P(X^2 \geq 26{,}124) = 0{,}001$
(ver a tabela)

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

n \ ε	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005	.001
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



$P(X^2 \leq 50) \sim 0{,}999$

Valor-p = $P(X^2 \geq 50) \sim 0{,}001$

Decisão:
Rejeita-se H0 para $\alpha = 1\%$, pois valor-p ~ 0,001 < 0,01. Existe evidência estatística para afirmar que o desvio padrão populacional é superior a 4 para $\alpha = 1\%$.

Testes de Hipóteses para σ_1^2/σ_2^2

Hipóteses, Estatística de Teste e Decisão

5

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

- **POPULAÇÕES NORMAIS**

Variância corrigida

$$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

Média	$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\dfrac{\bar{X}-\mu}{S'/\sqrt{n}} \sim t(n-1)$
Diferença de médias	$\dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} \sim N(0,1)$ $T=\dfrac{\dfrac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}}{\sqrt{\dfrac{(m-1)S_1'^2+(n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	$Z=\dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1'^2}{m}+\dfrac{S_2'^2}{n}}} \overset{a}{\sim} t(\nu)$ onde ν é o maior inteiro contido em r, $r=\dfrac{\left(\dfrac{s_1'^2}{m}+\dfrac{s_2'^2}{n}\right)^2}{\dfrac{1}{m-1}\left(\dfrac{s_1'^2}{m}\right)^2+\dfrac{1}{n-1}\left(\dfrac{s_2'^2}{n}\right)^2}$
Variância	$\dfrac{nS^2}{\sigma^2}=\dfrac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	
Relação de variâncias	$\dfrac{S_1'^2}{S_2'^2}\dfrac{\sigma_2^2}{\sigma_1^2} \sim F(m-1,n-1)$	

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.

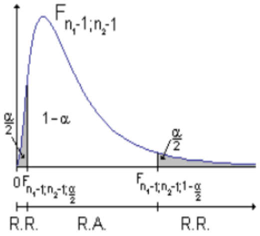
T. bilateral

Hipóteses a testar:

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

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

Regiões críticas:



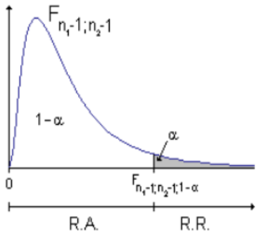
T. unilateral direito

Hipóteses a testar:

$$H_0: \sigma_1^2 \leq \sigma_0^2 \ vs: \sigma_1^2 > \sigma_2^2 \Leftrightarrow$$

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

Regiões críticas:



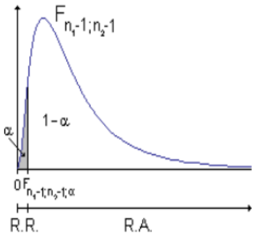
T. unilateral esquerdo

Hipóteses a testar:

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

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

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.

a) Com base num teste de hipóteses, ao nível de significância de 5%, pode concluir que a variabilidade é igual nos dois grupos?

b) Sem efectuar cálculos diga, justificando, qual a decisão que tomava no âmbito da alínea b), se considerasse um nível de significância de 1%?

c) Para o teste da alínea a) calcule o respectivo valor p e interprete.

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

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

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

ProbabilidadesEstatistica_2019 (uevora.pt)

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

ProbabilidadesEstatistica_2019 (uevora.pt)

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 α = 5% f_{obs} está na $R.A.$ então α = 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.

ProbabilidadesEstatistica_2019 (uevora.pt)

Obrigada!

Questões?