Estatística II

Licenciatura em Gestão

2.º Ano/1.º Semestre

2023/2024

Aulas Teóricas N.ºs 14 e 15 (Semana 8)

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 $\mu_1 - \mu_2$ (Variâncias Conhecidas)

Hipóteses, Estatística de Teste e Decisão (Amostras Independentes)

1

Teste t para a Diferença de Valores Médios (2 Amostras Independentes)

O **teste t para 2 amostras independentes** também é conhecido como teste *t* não-emparelhado

- Permite a **comparação de dois valores médios** usando amostras representativas de duas populações independentes.
- Os indivíduos são escolhidos **aleatoriamente** da população.
- As duas amostras são **independentes**.
- Deve-se saber se as **variâncias** são aproximadamente **iguais** (homocedasticidade) ou **não** (heterocedasticidade).
- **Suposição deste teste é**: A variável de interesse deve ter distribuição **Normal** em cada uma das populações (das quais as amostras foram recolhidas).

Nota: Geralmente, as variâncias são desconhecidas, mas também existem testes de hipóteses para o caso das variâncias serem conhecidas.

IC para μ$_1$-μ$_2$: Formulário

Variância corrigida

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

- **POPULAÇÕES NORMAIS**

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

Variâncias Conhecidas

Variâncias Desconhecidas e Iguais

Variâncias Desconhecidas e Diferentes

IC para μ_1-μ_2: Formulário

- **GRANDES AMOSTRAS**

Caso geral

Média	$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0,1)$	$\dfrac{\overline{X}-\mu}{S'/\sqrt{n}} \overset{a}{\sim} N(0,1)$
Diferença de médias	$\dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} \overset{a}{\sim} N(0,1)$	$\dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1'^2}{m}+\dfrac{S_2'^2}{n}}} \overset{a}{\sim} N(0,1)$

Variâncias Conhecidas

Variâncias Desconhecidas

Estatísticas de Teste t para a Diferença de Valores Médios (2 Amostras Independentes)

- **Variâncias iguais e desconhecidas:**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}$$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

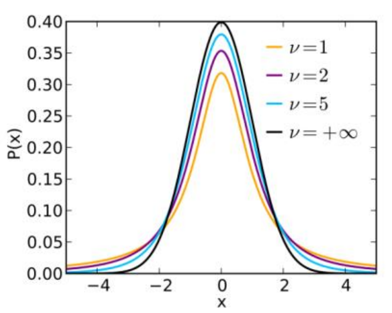
s: desvio padrão conjunto

- **Variâncias diferentes e desconhecidas:**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\nu = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{[(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)]}$$

Nota: Esta variável T tem distribuição t-Student com v graus de liberdade (g.l.´s). O valor dos g.l.´s é calculado através desta fórmula, sendo arredondado para baixo para o inteiro mais próximo.



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 variâncias 0,36 e 0,32, respectivamente.

a) Ao nível de significância de 1%, poderá afirmar que o ganho médio em peso das crianças alimentadas segundo o esquema A é:
   i. Igual ao das crianças alimentadas segundo o esquema B?
   ii. Superior ao das crianças alimentadas segundo o esquema B?
   iii. Inferior ao das crianças alimentadas segundo o esquema B?

b) A partir de que nível de significância rejeita cada uma das hipóteses anteriores?

ProbabilidadesEstatistica_2019 (uevora.pt)

Teste de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Conhecidas)

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 $X_1 \sim N\left(\mu_1 = ?\,;\, \sigma_1 = \sqrt{0{,}36}\right)$ e $X_2 \sim N\left(\mu_2 = ?;\, \sigma_2 = \sqrt{0{,}32}\right)$.

$n_1 = 9, \qquad \bar{x}_1 = 3{,}3367;$

$n_2 = 7, \qquad \bar{x}_2 = 3{,}6857.$

a) $\alpha = 1\%$.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (a) (i): Teste de Hipóteses para μ_1 - μ_2 (Variâncias Conhecidas)

i) $\mu_1 = \mu_2$?

Hipóteses:

$H_0: \mu_1 = \mu_2 \; vs \; H_1: \mu_1 \neq \mu_2$

$\Leftrightarrow H_0: \mu_1 - \mu_2 = 0 \; vs \; H_1: \mu_1 - \mu_2 \neq 0$ (teste bilateral).

Estatística de Teste:

Estatística de teste:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0;1).$$

$$z_{obs} = \frac{(3{,}3667 - 3{,}6857) - 0}{\sqrt{\frac{0{,}36}{9} + \frac{0{,}32}{7}}} = -1{,}0896.$$

Decisão (pela região de rejeição):

Pela tabela $z_{1-\frac{\alpha}{2}} = z_{0,995} = 2{,}576$.

Logo, $R.A.:]-2{,}576;\ 2{,}576[$ e $R.R.:]-\infty;\ -2{,}576] \cup [2{,}576; +\infty[$.

Como $z_{obs} \in R.A.$ não rejeitar H_0. Portanto, ao nível de significância de 1%, não existe evidência estatística de que o ganho médio em peso das crianças alimentadas segundo o esquema A seja diferente do das crianças alimentadas segundo o esquema B.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (a) (ii): Teste de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Conhecidas)

ii) $\mu_1 > \mu_2$?

Hipóteses:

$H_0: \mu_1 \leq \mu_2 \; vs. H_1: \mu_1 > \mu_2$
$\Leftrightarrow H_0: \mu_1 - \mu_2 \leq 0 \; vs. H_1: \mu_1 - \mu_2 > 0$ (teste unilateral direito).

Estatística de Teste:

Estatística de teste:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0;1).$$

Decisão (pela região de rejeição):

Como a estatística de teste é a mesma da alínea anterior, $z_{obs} = -1{,}0896$.

Pela tabela $z_{1-\alpha} = z_{0,99} = 2{,}326$. Logo, $R.A.:]-\infty;\ 2{,}326[$ e $R.R.: [2{,}326; +\infty[$.

Como $z_{obs} \in R.A.$ não rejeitar H_0. Portanto, ao nível de significância de 1%, não existe evidência estatística de que o ganho médio em peso das crianças alimentadas segundo o esquema A seja superior ao das crianças alimentadas segundo o esquema B.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (a) (iii): Teste de Hipóteses para μ_1 - μ_2 (Variâncias Conhecidas)

Hipóteses:

iii) $\mu_1 < \mu_2$?

$H_0: \mu_1 \geq \mu_2 \; vs \; H_1: \mu_1 < \mu_2$

$\Leftrightarrow H_0: \mu_1 - \mu_2 \geq 0 \; vs \; H_1: \mu_1 - \mu_2 < 0$ (teste unilateral esquerdo).

Estatística de Teste:

Estatística de teste: $Z = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0;1)$.

Decisão (pela região de rejeição):

Como a estatística de teste é a mesma da alínea anterior, $z_{obs} = -1{,}0896$.

Pela tabela $z_{1-\alpha} = z_{0,99} = 2{,}326$. Logo, $R.A.:]-2{,}326; +\infty[$ e $R.R.:]-\infty; -2{,}326]$.

Como $z_{obs} \in R.A.$ não rejeitar H_0. Portanto, ao nível de significância de 1%, não existe evidência estatística de que o ganho médio em peso das crianças alimentadas segundo o esquema A seja inferior ao das crianças alimentadas segundo o esquema B.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (b) (i) e (ii): Teste de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Conhecidas)

b) valor $p = P(\text{Rejeitar } H_0 | H_0 \text{ é verdadeira}) = P(Z_{obs} \in R.R. | \mu = \mu_0)$.

i) valor $p = 2 \times P(Z \geq |z_{obs}|) = 2 \times P(Z \geq 1{,}0896) = 2 \times \left(1 - \Phi(1{,}0896)\right)$
$= 2 \times (1 - 0{,}8621) = 0{,}2758$.

Decisão (pela valor-p):

A hipótese $H_0: \mu_1 = \mu_2$ é rejeitada para níveis de significância superiores ou iguais a 27,58%, logo não é rejeitada para qualquer nível de significância usual em investigação: não existe evidência estatística de que o ganho médio de peso das crianças alimentadas segundo o esquema A seja diferente do das crianças alimentadas segundo o esquema B.

ii) valor $p = P(Z \geq z_{obs}) = P(Z \geq -1{,}0896) = 1 - \Phi(-1{,}0896)$
$= 1 - \left(1 - \Phi(1{,}0896)\right) = \Phi(1{,}0896) = 0{,}8621$.

Decisão (pela valor-p):

Alternativa, com base no valor p bilateral calculado na alínea i): substituindo em H_1 μ_1 e μ_2 por $\overline{x}_1$ e $\overline{x}_2$, respectivamente, $\overline{x}_1 - \overline{x}_2 = 3{,}3367 - 3{,}6857 > 0$ dá uma proposição falsa. Logo,

$$\text{valor } p_{uni} = 1 - \frac{0{,}2758}{2} = 0{,}8621.$$

A hipótese $H_0: \mu_1 \leq \mu_2$ é rejeitada para níveis de significância superiores ou iguais a 86,21%. Assim, não existe evidência estatística de que o ganho médio de peso das crianças alimentadas segundo o esquema A seja superior ao das crianças alimentadas segundo o esquema B.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (b) (iii): Teste de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Conhecidas)

Decisão (pela valor-p):

iii) $valor\ p = P(Z \leq z_{obs}) = P(Z \leq -1{,}0896) = \Phi(-1{,}0896) = 1 - \Phi(1{,}0896)$
$= 1 - 0{,}8621 = 0{,}1379.$

Alternativa, com base no valor p bilateral calculado na alínea i): substituindo em H_1 μ_1 e μ_2 por $\overline{x}_1$ e $\overline{x}_2$, respectivamente, $\overline{x}_1 - \overline{x}_2 = 3{,}3367 - 3{,}6857 > 0$ dá uma proposição verdadeira. Logo,

$$\text{valor } p_{uni} = \frac{0{,}2758}{2} = 0{,}1379.$$

A hipótese H_0: $\mu_1 \geq \mu_2$ é rejeitada para níveis de significância superiores ou iguais a 13,79%: não existe evidência estatística de que o ganho médio de peso das crianças alimentadas segundo o esquema A seja inferior ao das crianças alimentadas segundo o esquema B.

ProbabilidadesEstatistica_2019 (uevora.pt)

Testes de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Desconhecidas e Iguais)

2

Hipóteses, Estatística de Teste e Decisão (Amostras Independentes)

Um determinado método de análise permite determinar o conteúdo de enxofre no petróleo bruto. Os ensaios efectuados em 10 e 8 amostras aleatórias de 1 kg de petróleo bruto, provenientes de furos pertencentes respectivamente aos campos A e B, revelaram os seguintes resultados (em gramas):

Campo A:	111	114	105	112	107	109	112	110	110	106
Campo B:	109	103	101	105	106	108	110	104		

a) Ao nível de significância de 10%, poderá afirmar que, em média, quantidade de enxofre por quilograma de petróleo do campo A é:

   i. Igual à do campo B?

   ii. Superior à do campo B?

   iii. Inferior à do campo B?

b) Calcule o $valor\ p$ associado a cada um dos testes anteriores.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (a) (i): Teste de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Desconhecidas e Iguais)

Sejam:

- X_1 a v.a. que representa a quantidade de enxofre por quilograma de petróleo do campo A,
- X_2 a v.a. que representa a quantidade de enxofre por quilograma de petróleo do campo B.

Nada é referido sobre a distribuição de X_1 e X_2.

$n_1 = 10$, $\overline{x}_1 = 109{,}6$ e $s_1 = 2{,}875$,
$n_2 = 8$, $\overline{x}_2 = 105{,}75$ e $s_2 = 3{,}105$.

a) $\alpha = 10\%$.

i) $\mu_1 = \mu_2$?

Hipóteses:

$H_0: \mu_1 = \mu_2 \ vs \ H_1: \mu_1 \neq \mu_2$
$\Leftrightarrow H_0: \mu_1 - \mu_2 = 0 \ vs \ H_1: \mu_1 - \mu_2 \neq 0$ (teste bilateral).

Para saber decidir qual a estatística de teste a utilizar, é preciso validar os pressupostos subjacentes:

- Normalidade:

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (a) (i): Teste de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Desconhecidas e Iguais)

- Igualdade das variâncias:

O I. C. a 90% para $\frac{\sigma_1^2}{\sigma_2^2}$ é dado por:

$$\left]\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1;n_2-1;1-\frac{\alpha}{2}}};\frac{S_1^2}{S_2^2}F_{n_2-1;n_1-1;1-\frac{\alpha}{2}}\right[.$$

Ver Tabela no Slide a seguir

Substituindo os valores, sendo $F_{9;\,7;0,95} = 3{,}68$ e $F_{7;\,9;0,95} = 3{,}29$, obtém-se:

$$\left]\frac{2{,}875^2}{3{,}105^2}\times\frac{1}{3{,}68};\frac{2{,}875^2}{3{,}105^2}\times 3{,}29\right[=]0{,}2332;2{,}8230[.$$

Como 1 pertence ao intervalo obtido, ao nível de significância de 10% não há evidências de que σ_1^2 seja diferente de σ_2^2. Portanto, pode-se considerar que $\sigma_1^2 = \sigma_2^2$.

ProbabilidadesEstatistica_2019 (uevora.pt)

Cálculo do Quantil da Distribuição F-Snedecor de Probabilidade $1-\alpha/2$ e com n1 e n2 g.l.´s

$$F_{m,n,\varepsilon} : P(X > F_{m,n,\varepsilon}) = \varepsilon$$

$F_{9;\,7;0,95} = 3{,}68$ e $F_{7;\,9;0,95} = 3{,}29$,

n – graus de liberdade do denominador		m - graus de liberdade do numerador																		
	ε	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
1	**.100**	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
	.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.90	245.95	248.02	249.05	250.10	251.14	252.20	253.25	254.32
	.025	647.79	799.48	864.15	899.60	921.83	937.11	948.20	956.64	963.28	968.63	976.72	984.87	993.08	997.27	1001.40	1005.60	1009.79	1014.04	1018.26
	.010	4052.18	4999.34	5403.53	5624.26	5763.96	5858.95	5928.33	5980.95	6022.40	6055.93	6106.68	6156.97	6208.66	6234.27	6260.35	6286.43	6312.97	6339.51	6365.59
2	**.100**	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
	.010	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.48	99.48	99.49	99.50
3	**.100**	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	**.100**	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	**.100**	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.11
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	**.100**	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	**.100**	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.41	4.36	4.31	4.25	4.20	4.14
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	**.100**	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
	.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
	.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
	.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	**.100**	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
	.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
	.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
	.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31

Exercício (a) (i): Teste de Hipóteses para μ_1 - μ_2 (Variâncias Desconhecidas e Iguais)

Estatística de Teste:

Deste modo, a estatística de teste a usar é:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2=10+8-2=16}.$$

$$t_{obs} = \frac{(109{,}6 - 105{,}75) - 0}{\sqrt{\frac{(10-1)2{,}875 + (8-1)3{,}105}{10+8-2}}\sqrt{\frac{1}{10}+\frac{1}{8}}} = 2{,}7255.$$

Decisão (pela região de rejeição):

Pela tabela $t_{n_1+n_2-2;\,1-\frac{\alpha}{2}} = t_{16;\,0{,}95} = 1{,}746.$

Logo, $R.A.:]-1{,}746;\ 1{,}746[$ e $R.R.:]-\infty;\ -1{,}746] \cup [1{,}746; +\infty[$.

Como $t_{obs} \in R.R.$ rejeitar H_0. Portanto, ao nível de significância de 10%, existe evidência estatística de que, em média, a quantidade de enxofre por quilograma de petróleo do campo A é diferente da do campo B.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (a) (ii): Teste de Hipóteses para μ1 - μ2 (Variâncias Desconhecidas e Iguais)

ii) $\mu_1 > \mu_2$?

Hipóteses:

$H_0: \mu_1 \leq \mu_2 \ vs \ H_1: \mu_1 > \mu_2$

$\Leftrightarrow H_0: \mu_1 - \mu_2 \leq 0 \ vs \ H_1: \mu_1 - \mu_2 > 0$ (teste unilateral direito).

Estatística de Teste:

Estatística de teste:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2=10+8-2=16}.$$

Como a estatística de teste é a mesma da alínea anterior, $t_{obs} = 2{,}7255$.

Decisão (pela região de rejeição):

Pela tabela $t_{n_1+n_2-2;\,1-\alpha} = t_{16;\,0{,}90} = 1{,}337$.

Logo, $R.A.:]-\infty;\ 1{,}337[$ e $R.R.: [1{,}337; +\infty[$.

Como $t_{obs} \in R.R.$ rejeitar H_0. Portanto, ao nível de significância de 10%, existe evidência estatística de que, em média, a quantidade de enxofre por quilograma de petróleo do campo A é superior à do campo B.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (a) (iii): Teste de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Desconhecidas e Iguais)

iii) $\mu_1 < \mu_2$?

Hipóteses:

$H_0: \mu_1 \geq \mu_2 \; vs \; H_1: \mu_1 < \mu_2$

$\Leftrightarrow H_0: \mu_1 - \mu_2 \geq 0 \; vs \; H_1: \mu_1 - \mu_2 < 0$ (teste unilateral esquerdo).

Estatística de Teste:

Estatística de teste:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2=10+8-2=16}.$$

Como a estatística de teste é a mesma da alínea anterior, $t_{obs} = 2{,}7255$.

Decisão (pela região de rejeição):

Pela tabela $t_{n_1+n_2-2;\,1-\alpha} = t_{16;\,0{,}90} = 1{,}337$.
Logo, $R.A.:]-1{,}337; +\infty[$ e $R.R.:]-\infty; -1{,}337]$.

Como $t_{obs} \in R.A.$ não rejeitar H_0. Portanto, ao nível de significância de 10%, não existe evidência estatística de que, em média, a quantidade de enxofre por quilograma de petróleo do campo A seja inferior à do campo B.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (b): Teste de Hipóteses para μ_1 - μ_2 (Variâncias Desconhecidas e Iguais)

Decisão (pela valor-p):

b) valor $p = P(\text{Rejeitar } H_0 | H_0 \text{ é verdadeira}) = P(Z_{obs} \in R.R. | \mu = \mu_0)$.

i) valor $p = 2 \times P(T \geq |t_{obs}|) = 2 \times P(T \geq 2{,}7255) = 2 \times \left(1 - P(T < 2{,}7255)\right)$
$= 2 \times (1 - 0{,}9925) = 0{,}015$.

A hipótese $H_0: \mu_1 = \mu_2$ é rejeitada para níveis de significância superiores ou iguais a 1,5%. Logo, para um nível de significância de 5% existe evidência de que o teor médio de enxofre no campo A é diferente do campo B, mas para 1% essa afirmação já não pode ser sustentada. Repare-se que o *valor p* calculado é o valor indicado no quadro de resultados do SPSS anteriormente apresentado.

ii) valor $p = P(T \geq t_{obs}) = P(T \geq 2{,}7255) = 1 - P(T < 2{,}7255) = 1 - 0{,}9925 = 0{,}0075$.
Alternativa, com base no *valor-p* bilateral calculado na alínea i): substituindo em H_1 μ_1 e μ_2 por $\overline{x}_1$ e $\overline{x}_2$, respectivamente, $\overline{x}_1 - \overline{x}_2 = 109{,}6 - 105{,}75 > 0$ dá uma proposição verdadeira. Logo,

$$valor\text{-}p_{uni} = \frac{0{,}015}{2} = 0{,}0075.$$

A hipótese $H_0: \mu_1 \leq \mu_2$ é rejeitada para níveis de significância superiores ou iguais a 0,75%. Assim, para qualquer nível de significância sensato/usual ($\leq 10\%$) existe evidência de que o teor médio de enxofre no campo A é superior ao do campo B.

iii) valor $p = P(T \leq t_{obs}) = P(T \leq 2{,}7255) = 0{,}9925$.
Alternativa, com base no *valor-p* bilateral calculado na alínea i): substituindo em H_1 μ_1 e μ_2 por $\overline{x}_1$ e $\overline{x}_2$, respectivamente, $\overline{x}_1 - \overline{x}_2 = 109{,}6 - 105{,}75 > 0$ dá uma proposição falsa. Logo,

$$\text{valor } p_{uni} = 1 - \frac{0{,}015}{2} = 0{,}9925.$$

A hipótese $H_0: \mu_1 \geq \mu_2$ é rejeitada para níveis de significância superiores ou iguais a 99,25. Assim, existe não existe evidência de que o teor médio de enxofre no campo A seja inferior ao do campo B.

ProbabilidadesEstatistica_2019 (uevora.pt)

Testes de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Desconhecidas e Diferentes)

3

Hipóteses, Estatística de Teste e Decisão
(Amostras Independentes)

Para um estudo sobre a caracterização da altura da população portuguesa, foi recolhida uma amostra de 1861 pessoas, com as seguintes características: (conjunto de dados semelhante ao disponibilizado no Capítulo 12, mas com uma amostra de maior dimensão):

Group Statistics

	Sexo	N	Mean	Std. Deviation
Altura	Masculino	853	168,46	7,617
	Feminino	1007	158,48	6,652

Supondo a Normalidade das distribuições e assumindo que as variâncias populacionais são desconhecidas e diferentes, verifique se se pode considerar que as alturas médias dos homens e das mulheres são iguais, com 95% de confiança.

a) Suspeita que em média a altura dos homens não é igual à das mulheres. Teste esta hipótese ao nível de significância de 5%.
b) Calcule o valor p associado ao teste da alínea anterior.
c) Teste a hipótese de a média da altura dos homens ser superior à das mulheres, ao nível de significância de 0,5%?
d) Determine o valor p associado ao teste anterior.
e) Ao nível de significância de 2,5%, pode-se afirmar que em média a altura dos homens é superior à das mulheres?
f) A partir de que nível de significância é rejeitada a hipótese do teste anterior?

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (a): Teste de Hipóteses para μ_1 - μ_2 (Variâncias Desconhecidas e Diferentes)

Sejam:

- X_1 a v.a. que representa a altura dos homens,
- X_2 a v.a. que representa a altura das mulheres,

com $X_1 \sim N(\mu_1 =?\,; \sigma_1 =?)$ e $X_2 \sim N(\mu_2 = ?; \sigma_2 =?)$, mas $\sigma_1^2 \neq \sigma_2^2$.

$n_1 = 853,\quad \overline{x}_1 = 168{,}46$ e $s_1 = 7{,}617$,
$n_2 = 1007,\quad \overline{x}_2 = 158{,}48$ e $s_2 = 6{,}652$.

a) $\alpha = 5\%$, $\mu_1 \neq \mu_2$?

Hipóteses:

$H_0{:}\, \mu_1 = \mu_2 \; vs H_1{:}\, \mu_1 \neq \mu_2$
$\Leftrightarrow H_0{:}\, \mu_1 - \mu_2 = 0 \; vs \; H_1{:}\, \mu_1 - \mu_2 \neq 0$ (teste bilateral).

Estatística de Teste:

Estatística de teste:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{\circ}{\sim} t_v, \text{ onde } v = \left[\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{S_2^2}{n_2}\right)^2} \right].$$

$$t_{obs} = \frac{(168{,}46 - 158{,}48) - 0}{\sqrt{\frac{7{,}617^2}{853} + \frac{6{,}652^2}{1007}}} = 29{,}816.$$

$$v = \left[\frac{\left(\frac{7{,}617^2}{853} + \frac{6{,}652^2}{1007}\right)^2}{\frac{1}{853 - 1}\left(\frac{7{,}617^2}{853}\right)^2 + \frac{1}{1007 - 1}\left(\frac{6{,}652^2}{1007}\right)^2} \right] = [1705{,}6] = 1705.$$

Decisão (pela região de rejeição):

Pela tabela $t_{v;\,1-\frac{\alpha}{2}} = t_{1705\,;\,0{,}975} = 1{,}96$.

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

Como $t_{obs} \in R.R.$ rejeitar H_0. Portanto, ao nível de significância de 5%, existe evidência estatística de existe diferença significativa entre as médias das alturas dos homens e das mulheres.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (b): Teste de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Desconhecidas e Diferentes)

Decisão (pela valor-p):

b) valor $p = 2 \times P(T \geq |t_{obs}|) = 2 \times P(T \geq 29{,}816) = 2 \times \left(1 - P(T < 29{,}816)\right)$

$$\approx 2 \times (1 - 1) = 0.$$

Logo, aos níveis usuais de significância existe evidência de que a média das alturas dos homens difere das mulheres. Repare-se que o valor p calculado é o valor indicado no quadro de resultados do SPSS anteriormente apresentado.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (c): Teste de Hipóteses para μ_1 - μ_2 (Variâncias Desconhecidas e Diferentes)

Hipóteses:

c) $\alpha = 0{,}5\%, \mu_1 > \mu_2$?

$$H_0: \mu_1 \leq \mu_2 \ vs \ H_1: \mu_1 > \mu_2$$

$$\Leftrightarrow H_0: \mu_1 - \mu_2 \leq 0 \ vs \ H_1: \mu_1 - \mu_2 > 0 \text{ (teste unilateral direito).}$$

Estatística de Teste:

Estatística de teste:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{\circ}{\sim} t_v, \text{ onde } v = \left[\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{S_2^2}{n_2}\right)^2}\right].$$

Como a estatística de teste é a mesma da alínea anterior, $t_{obs} = 29{,}816$ e $v = 1705$.

Decisão (pela região de rejeição):

Pela tabela $t_{v;\,1-\alpha} = t_{1705\,;\,0{,}995} = 2{,}576$.
Logo, $R.A.:]-\infty;\ 2{,}576[$ e $R.R.: [2{,}576; +\infty[$.

Como $t_{obs} \in R.R.$ rejeitar H_0. Portanto, ao nível de significância de 0,5%, existe evidência estatística de que, em média, os homens são mais altos do que as mulheres.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (d): Teste de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Desconhecidas e Diferentes)

Decisão (pela valor-p):

d) valor $p = P(T \geq t_{obs}) = P(T \geq 29{,}816) = 1 - P(T < 29{,}816) \approx 1 - 1 = 0$

Alternativa, com base no *valor-p* bilateral calculado na alínea i): substituindo em H_1 μ_1 e μ_2 por $\overline{x}_1$ e $\overline{x}_2$, respectivamente, $\overline{x}_1 - \overline{x}_2 = 168{,}46 - 158{,}48 > 0$ dá uma proposição verdadeira. Logo,

$$\text{valor } p_{uni} \approx \frac{0}{2} = 0.$$

A hipótese $H_0: \mu_1 \leq \mu_2$ é rejeitada para níveis de significância de aproximadamente 0. Assim, para qualquer nível de significância sensato/usual ($\leq 10\%$) existe evidência de que em média os homens são mais altos do que as mulheres.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (e): Teste de Hipóteses para $\mu_1 - \mu_2$ (Variâncias Desconhecidas e Diferentes)

Hipóteses:

e) $\alpha = 2{,}5\%, \mu_1 < \mu_2$?

$$H_0: \mu_1 \geq \mu_2 \; vs \; H_1: \mu_1 < \mu_2$$

$$\Leftrightarrow H_0: \mu_1 - \mu_2 \geq 0 \; vs \; H_1: \mu_1 - \mu_2 < 0 \text{ (teste unilateral esquerdo).}$$

Estatística de Teste:

Estatística de teste:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{\circ}{\sim} t_v, \text{ onde } v = \left[\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{S_2^2}{n_2}\right)^2} \right].$$

Como a estatística de teste é a mesma da alínea anterior, $t_{obs} = 29{,}816$ e $v = 1705$.

Decisão (pela região de rejeição):

Pela tabela $t_{v;\,1-\alpha} = t_{1705\,;\,0{,}975} = 1{,}96$.
Logo, $R.A.:]-1{,}96; +\infty[$ e $R.R.:]-\infty; -1{,}96]$.

Como $t_{obs} \in R.A.$ não rejeitar H_0. Portanto, ao nível de significância de 2,5%, não existe evidência estatística de que, em média, a altura dos homens seja inferior à das mulheres.

ProbabilidadesEstatistica_2019 (uevora.pt)

Exercício (f): Teste de Hipóteses para μ_1 - μ_2 (Variâncias Desconhecidas e Diferentes)

Decisão (pela valor-p):

f) valor $p = P(T \leq t_{obs}) = P(T \leq 29{,}816\ ) \approx 1$.
Alternativa, com base no *valor-p* bilateral calculado na alínea i): substituindo em H_1 μ_1 e μ_2 por $\overline{x}_1$ e $\overline{x}_2$, respectivamente, $\overline{x}_1 - \overline{x}_2 = 168{,}46 - 158{,}48 > 0$ dá uma proposição falsa. Logo,

$$\text{valor } p_{uni} \approx 1 - \frac{0}{2} = 1.$$

A hipótese H_0: $\mu_1 \geq \mu_2$ é rejeitada para níveis de significância de aproximadamente 1. Assim, para qualquer nível de significância sensato/usual ($\leq 10\%$) não existe evidência de que em média a altura dos homens seja inferior à das mulheres.

ProbabilidadesEstatistica_2019 (uevora.pt)

Obrigada!

Questões?