

1º semestre – 2011/2012  
 7/01/2012 – **10:30**

2º Teste (Época Normal)  
 Duração: 1 hora e 30 minutos

**Grupo I**

**1.5 + 1.5 + 3.5 + 3.5 = 10.0** valores

**Exercício 1**

- (a) • V.a. de interesse

$$X$$

- A.a.

$X = (X_1, \dots, X_n)$  a.a. de dimensão  $n$  proveniente da população  $X$   
 $X_i \stackrel{i.i.d.}{\sim} X, i = 1, \dots, n$

- Parâmetros DESCONHECIDOS

$$E(X_i) = E(X) = \mu$$

$$V(X_i) = V(X) = \sigma^2$$

- V.a. auxiliar

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

- Resultados auxiliares

$$E(Z) = n - 1$$

$$V(Z) = 2(n - 1)$$

- Estimador de  $\sigma^2$

$$T = c \sum_{i=1}^n (X_i - \bar{X})^2 = c\sigma^2 Z$$

- Erro quadrático médio de  $T$

Ao tirar-se partido desta fórmula alternativa de  $T$  e dos resultados auxiliares, tem-se:

$$\begin{aligned} EQM_{\sigma^2}(T) &= E[(T - \sigma^2)^2] \\ &= V(T) + [E(T) - \sigma^2]^2 \\ &= V(c\sigma^2 Z) + [E(c\sigma^2 Z) - \sigma^2]^2 \\ &= c^2 \sigma^4 V(Z) + [c\sigma^2 E(Z) - \sigma^2]^2 \\ &= c^2 \sigma^4 \times 2(n - 1) + [c\sigma^2(n - 1) - \sigma^2]^2 \\ &= 2(n - 1)c^2 \sigma^4 + [c(n - 1) - 1]^2 \sigma^4. \end{aligned}$$

- (b) • Minimização de  $EQM_{\sigma^2}(T)$

Atendendo ao facto de o ponto de mínimo de  $EQM_{\sigma^2}(T) = 2(n - 1)c^2 \sigma^4 + [c(n - 1) - 1]^2 \sigma^4$  coincidir com o da função  $2(n - 1)c^2 + [c(n - 1) - 1]^2$ , pretende identificar-se

$$c^* : \begin{cases} \frac{dEQM_{\sigma^2}(T)}{dc} \Big|_{c=c^*} = 0 & \text{(ponto de estacionaridade)} \\ \frac{d^2EQM_{\sigma^2}(T)}{dc^2} \Big|_{c=c^*} > 0 & \text{(ponto de mínimo)} \end{cases}$$

$$\begin{cases} 4(n - 1)c^* + 2(n - 1)[c^*(n - 1) - 1] = 0 \\ 4(n - 1) + 2(n - 1)^2 > 0 \end{cases}$$

$$\begin{cases} 2c^* + [c^*(n - 1) - 1] = 0 \Leftrightarrow c^* + nc^* = 1 \Leftrightarrow c^* = \frac{1}{n+1} \\ \text{Proposição verdadeira para qualquer inteiro positivo } n \end{cases}$$

- **Obs.**

$EQM_{\sigma^2}(T)$  é efectivamente mínimo quando  $c = \frac{1}{n+1}$ . Assim se conclui que, no contexto da estimação da variância de uma população normal com valor esperado e variância desconhecidos, nem a variância corrigida da amostra aleatória,  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ , nem o estimador de máxima verosimilhança de  $\sigma^2$ ,  $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ , minimizam o EQM.

## Exercício 2

- (a)    • **V.a. de interesse**

$X$  = eficiência térmica de motores Diesel

- **Situação**

$X \sim \text{Normal}(\mu, \sigma^2)$

$\mu$  DESCONHECIDO

$\sigma^2$  desconhecido

- **Hipóteses**

$$H_0 : \mu = \mu_0 = 35$$

$$H_1 : \mu < \mu_0 = 35$$

- **Estatística de teste**

$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \sim_{H_0} t_{(n-1)}$$

pois pretendemos efectuar um teste sobre o valor esperado de uma população normal com variância desconhecida.

- **Região de rejeição de  $H_0$**  (para valores da estatística de teste)

Tratando-se de um teste unilateral inferior ( $H_1 : \mu < \mu_0$ ), a região de rejeição de  $H_0$  é um intervalo à esquerda do tipo  $W = (-\infty, c)$ .

- **Decisão**

Uma vez que

$$n = 25$$

$$\bar{x} = 31.4$$

$$s = 1.6$$

o valor observado da estatística é igual a

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \\ &= \frac{31.4 - 35}{\frac{1.6}{\sqrt{25}}} \\ &= -11.25 \end{aligned}$$

### Decisão (cont.) e intervalo para o p-value

Dado que a região de rejeição deste teste é um intervalo à esquerda, tem-se:

$$\begin{aligned} p-value &= P(T < t \mid H_0) \\ &= P(T < -11.25 \mid H_0) \\ &= F_{t_{(25-1)}}(-11.25) \\ &= 1 - F_{t_{(24)}}(11.25) \end{aligned}$$

Recorrendo às tabelas de quantis da distribuição de t-student podemos adiantar um intervalo para o *p-value*. Efectivamente,

$$F_{t_{(24)}}^{-1}(0.9995) = 3.745 < 11.25$$

$$0.9995 < F_{t_{(24)}}(11.25)$$

$$0.0005 = 1 - 0.9995 > 1 - F_{t_{(24)}}(11.25) = p-value.$$

Deste modo, o intervalo para o *p-value* é  $(0, 0.0005)$ , pelo que devemos rejeitar  $H_0$  a qualquer n.s.  $\alpha_0 \geq 0.05\%$ , nomeadamente aos níveis usuais de significância de 1%, 5% e 10%.

- **Decisão (cont.) e obtenção do p-value usando máquina de calcular**

Uma vez que a região de rejeição deste teste é um intervalo à esquerda, tem-se:

$$\begin{aligned} p\text{-value} &= P(T < t \mid H_0) \\ &= P(T < -11.25 \mid H_0) \\ &= F_{t_{(25-1)}}(-11.25) \\ &= 2.34858 \times 10^{-11} \end{aligned}$$

pelo que

- não devemos rejeitar  $H_0$  a qualquer n.s.  $\alpha_0 \leq 2.34858 \times 10^{-9}\%$ ;
- rejeitar  $H_0$  a qualquer n.s.  $\alpha_0 > 2.34858 \times 10^{-9}\%$ , nomeadamente aos níveis usuais de significância de 1%, 5% e 10%.

(b) • **V.a. de interesse**

$X$  = eficiência térmica de motores Diesel

- **Situação**

$$X \sim \text{Normal}(\mu, \sigma^2)$$

$\mu$  DESCONHECIDO

$\sigma^2$  desconhecido

- **Passo 1 — Selecção da v.a. fulcral para  $\sigma^2$**

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

uma vez que é suposto determinar um IC para a variância de uma população normal, com valor esperado desconhecido.

- **Passo 2 — Obtenção dos quantis de probabilidade**

Ao ter-se em consideração que  $n = 25$  e  $(1 - \alpha) \times 100\% = 99\%$  far-se-á uso dos quantis não simétricos:

$$\begin{cases} a_\alpha = F_{\chi^2_{(n-1)}}^{-1}(\alpha/2) = F_{\chi^2_{(24)}}^{-1}(0.01) \stackrel{\text{tabela}}{=} 9.886 \\ b_\alpha = F_{\chi^2_{(n-1)}}^{-1}(1 - \alpha/2) = F_{\chi^2_{(24)}}^{-1}(0.99) \stackrel{\text{tabela}}{=} 45.56. \end{cases}$$

- **Passo 3 — Inversão da desigualdade  $a_\alpha \leq Z \leq b_\alpha$**

$$P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha$$

$$P\left[a_\alpha \leq \frac{(n-1)S^2}{\sigma^2} \leq b_\alpha\right] = 1 - \alpha$$

$$P\left[\frac{1}{b_\alpha} \leq \frac{\sigma^2}{(n-1)S^2} \leq \frac{1}{a_\alpha}\right] = 1 - \alpha$$

$$P\left[\sqrt{\frac{(n-1)S^2}{b_\alpha}} \leq \sigma \leq \sqrt{\frac{(n-1)S^2}{a_\alpha}}\right] = 1 - \alpha$$

- **Passo 4 — Concretização**

Atendendo ao par de quantis acima e ao facto de

$$n = 25$$

$$s^2 = 1.6^2$$

e

$$IC_{(1-\alpha) \times 100\%}(\sigma) = \left[ \sqrt{\frac{(n-1)s^2}{F_{\chi^2_{(n-1)}}^{-1}(1-\alpha/2)}}, \sqrt{\frac{(n-1)s^2}{F_{\chi^2_{(n-1)}}^{-1}(\alpha/2)}} \right],$$

segue-se:

$$\begin{aligned}
 IC_{99\%}(\sigma) &= \left[ \sqrt{\frac{(25-1) \times 1.6^2}{45.56}}, \sqrt{\frac{(24-1) \times 1.6^2}{9.886}} \right] \\
 &\simeq [\sqrt{1.348551}, \sqrt{6.214849}] \\
 &= [1.161271, 2.492960].
 \end{aligned}$$

## Grupo II

**4.0 + 3.0 + 3.0 = 10.0** valores

### Exercício 1

- (a) • **V.a. de interesse**

$X$  = tempo até conclusão do teste (em minutos)

- **Hipóteses**

$$H_0 : X \sim \text{Normal}(\mu = 100, \sigma^2 = 15.6^2)$$

$$H_1 : X \not\sim \text{Normal}(\mu = 100, \sigma^2 = 15.6^2)$$

- **Nível de significância**

$$\alpha_0 = 5\%$$

- **Estatística de Teste**

$$T = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \stackrel{a}{\sim}_{H_0} \chi^2_{(k-\beta-1)},$$

onde:

$$k = \text{No. de classes} = 4$$

$O_i$  = Frequência absoluta observável da classe  $i$

$E_i$  = Frequência absoluta esperada, sob  $H_0$ , da classe  $i$

$$\beta = \text{No. de parâmetros a estimar} = 0.^1$$

- **Região de rejeição de  $H_0$**  (para valores de  $T$ )

A região de rejeição de  $H_0$  escrita para valores de  $T$  é um intervalo à direita  $W = (c, +\infty)$ , onde

$$c = F_{\chi^2_{(k-\beta-1)}}^{-1}(1 - \alpha_0) = F_{\chi^2_{(4-0-1)}}^{-1}(1 - 0.05) = F_{\chi^2_{(3)}}^{-1}(0.95) \stackrel{\text{tabela}}{=} 7.815.$$

- **Cálculo das frequências absolutas esperadas sob  $H_0$**

Tendo em conta que  $P(X \leq x | H_0) = \Phi\left(\frac{x-100}{15.6}\right)$ , as probabilidades de pertença às 4 classes são, sob  $H_0$ , iguais a:

$$\begin{aligned}
 p_1^0 &= P[X \leq 90 | H_0] \\
 &= \Phi\left(\frac{90-100}{15.6}\right) \\
 &\simeq \Phi(-0.64) \\
 &= 1 - \Phi(0.64) \\
 &\stackrel{\text{tabela}}{=} 1 - 0.7389 \\
 &= 0.2611 \\
 p_2^0 &= P[90 < X \leq 100 | H_0] \\
 &= \Phi\left(\frac{100-100}{15.6}\right) - \Phi\left(\frac{90-100}{15.6}\right) \\
 &\simeq \Phi(0) - \Phi(-0.64) \\
 &= 0.5 - 0.2611 \\
 &= 0.2389
 \end{aligned}$$

<sup>1</sup>Dado que a distribuição conjecturada em  $H_0$  está completamente especificada, i.e.,  $H_0$  é uma hipótese simples.

$$\begin{aligned}
p_3^0 &= P[100 < X \leq 110 \mid H_0] \\
&= \Phi\left(\frac{110 - 100}{15.6}\right) - \Phi\left(\frac{100 - 100}{15.6}\right) \\
&\simeq \Phi(0.64) - \Phi(0) \\
&= 0.7389 - 0.5 \\
&= 0.2389 \\
p_4^0 &= P[X > 110 \mid H_0] \\
&= 1 - \Phi\left(\frac{110 - 100}{15.6}\right) \\
&\simeq 1 - \Phi(0.64) \\
&= 1 - 0.7389 \\
&= 0.2611.
\end{aligned}$$

As frequências absolutas esperadas sob  $H_0$  são dadas por  $E_i = n \times p_i^0 = 100 \times p_i^0$ ,  $i = 1, 2, 3, 4$ ,  
<sup>2</sup> logo iguais a:  $E_1 = 26.11$ ,  $E_2 = 23.89$ ,  $E_3 = 23.89$  e  $E_4 = 26.11$ .<sup>3</sup>

#### • Decisão

No cálculo do valor observado da estatística de teste convém adiantar a seguinte tabela auxiliar:

$i$	Classe $i$	Freq. abs. obs. $o_i$	Freq. abs. esper. sob $H_0$ $E_i = n \times p_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i - E_i)^2}{E_i}$
1	$(-\infty, 90]$	29	26.11	$\frac{(29 - 26.11)^2}{26.11} = 0.319881$
2	$(90, 100]$	21	23.89	0.349607
3	$(100, 110]$	25	23.89	0.051574
4	$(110, +\infty)$	25	26.11	0.047189
$\sum_{i=1}^k o_i = n = 100$		$\sum_{i=1}^k E_i = n = 100$	$t = \sum_{i=1}^k \frac{(o_i - E_i)^2}{E_i} = 0.768251$	

Como  $t = 0.768251 \notin W = (7.815, +\infty)$  não devemos rejeitar a hipótese de os dados serem provenientes de uma população com distribuição Normal( $\mu = 100$ ,  $\sigma^2 = 15.6^2$ ), a qualquer nível de significância menor ou igual que 5%.

### Exercício 2

(a) • [Modelo de RLS]

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$Y_i$  = tempo até fractura da  $i$  – ésima peça

$x_i$  = pressão aplicada à  $i$  – ésima peça

$\epsilon_i$  = erro aleatório associado à medição do tempo até fractura da  $i$  – ésima peça

• Hipóteses de trabalho

$\epsilon_i \sim_{i.i.d.} \text{Normal}(0, \sigma^2)$ ,  $i = 1, \dots, n$  (hipótese de trabalho)

$\beta_0, \beta_1, \sigma^2$  DESCONHECIDOS]

• Estimativas de  $\beta_0$  e  $\beta_1$

Dado que

$$n = 7$$

$$\sum_{i=1}^n x_i = 569$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{569}{7} = 81.285714$$

<sup>2</sup>Importa que não será necessário qualquer agrupamento de classes pois em pelo menos 80% das classes verifica-se  $E_i \geq 5$ , como veremos já de seguida.

<sup>3</sup>Notar que não é necessário qualquer agrupamento de classes uma vez que  $E_i \geq 1$ ,  $\forall i$  e  $E_i \geq 5$  em pelo menos 80% das classes.

$$\begin{aligned}
\sum_{i=1}^n x_i^2 &= 46375 \\
\sum_{i=1}^n x_i^2 - n(\bar{x})^2 &= 46375 - 7 \times 81.285714^2 = 123.428571 \\
\sum_{i=1}^n y_i &= 6310 \\
\bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i = \frac{6310}{7} = 901.428571 \\
\sum_{i=1}^n y_i^2 &= 5764600 \\
\sum_{i=1}^n y_i^2 - n(\bar{y})^2 &= 5764600 - 7 \times 901.428571^2 = 76585.714286
\end{aligned}$$

$$\sum_{i=1}^n x_i y_i = 510410$$

$$\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} = 510410 - 7 \times 81.285714 \times 901.428571 = -2502.857143,$$

as estimativas de  $\beta_1$  e  $\beta_0$  são, para este modelo, iguais a:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2} \\
&= \frac{-2502.857143}{123.428571} \\
&= -20.2777778 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&= 901.428571 - (-20.2777778) \times 81.285714 \\
&= 2549.722222
\end{aligned}$$

- **Recta de regressão**

$$\hat{E}(Y|x) = \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times x = 2549.722222 - 20.2777778 x$$

- **Estimativa de  $\sigma^2$**

Atendendo ao facto de  $V(\epsilon) = \sigma^2$  e ao formulário, tem-se:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2} \left[ \left( \sum_{i=1}^n y_i^2 - n \bar{y}^2 \right) - (\hat{\beta}_1)^2 \left( \sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) \right] \\
&= \frac{1}{7-2} (76585.714286 - (-20.2777778)^2 \times 123.428571) \\
&= 5166.666667.
\end{aligned}$$

(b)

- **Hipóteses**

$$H_0 : \beta_1 = \beta_{1,0} = -20 \text{ vs. } H_1 : \beta_1 \neq \beta_{1,0}$$

- **Nível de significância**

$$\alpha_0 = 10\%$$

- **Estatística de teste**

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}}} \sim_{H_0} t_{(n-2)}$$

- **Região de rejeição de  $H_0$**  (para valores da estatística de teste)

Estamos a lidar com um teste bilateral ( $H_1 : \beta_1 \neq \beta_{1,0}$ ), pelo que a região de rejeição de  $H_0$  é  $W = (-\infty, -c) \cup (c, +\infty)$ , onde

$$\begin{aligned}
c &: P(\text{Rejeitar } H_0 | H_0) = \alpha_0 \\
c &= F_{t_{(n-2)}}^{-1} \left( 1 - \frac{\alpha_0}{2} \right) \\
c &= F_{t_{(5)}}^{-1} (0.95) \\
c &= 2.015
\end{aligned}$$

- **Decisão**

Tendo em conta que

$$\hat{\beta}_1 = -20.2777778$$

$$\hat{\sigma}^2 = 5166.666667$$

$$\sum_{i=1}^n x_i^2 - n \bar{x}^2 = 123.428571,$$

o valor observado da estatística de teste é dado por

$$\begin{aligned} t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}}} \\ &= \frac{-20.2777778 - (-20)}{\sqrt{\frac{5166.666667}{123.428571}}} \\ &= -0.042933 \end{aligned}$$

Como  $t = -0.042934 \notin W = (-\infty, -2.015) \cup (2.015, +\infty)$ , não devemos rejeitar  $H_0$  a qualquer n.s.  $\alpha_0 \leq 10\%$ , nomeadamente a qualquer nível usual de significância de 1%, 5%, 10%.