

Probabilidades e Estatística

LEAN, LEGM, LEIC-A, LEMat, LMAC, MEAer, MEAmbi,
MEBiol, MEBiom, MEC, MEEC, MEFT, MEMec MEQ

1º semestre – 2011/2012
2/02/2012 – 11:30

2º TESTE (Época Recurso)
Duração: 1 hora e 30 minutos

Grupo I

2.0 + 4.0 + 4.0 = 10.0 valores

Exercício 1

- Estimador de θ e suas propriedades

T é tal que $E(T) = \theta$ e $V(T) > 0$.

- Estimador de θ^2

T^2

- Estimador centrado de θ^2 ?

T^2 é um estimador centrado de θ^2 sse $E(T^2) = \theta^2, \forall \theta$. Ora, como $V(T) > 0$ tem-se

$$\begin{aligned} E(T^2) &= V(T) + E^2(T) \\ &> 0 + \theta^2 \\ &= \theta^2, \end{aligned}$$

para qualquer θ . Assim, conclui-se que $E(T^2) \neq \theta^2$, logo T^2 não é um estimador centrado de θ^2 .

Exercício 2

- V.a. de interesse

X = número de defeitos em cada produto

- Distribuição

$X \sim \text{Poisson}(\lambda)$

- Parâmetro DESCONHECIDO

$\lambda, \lambda > 0$

- F.p.

$P(X = x) \stackrel{\text{form}}{=} \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, \dots$

- Amostra

$\underline{x} = (x_1, \dots, x_n)$ amostra de dimensão n proveniente da população X

- Obtenção da estimativa de MV de λ

Passo 1 — Função de verosimilhança

$$\begin{aligned} L(\lambda|\underline{x}) &\stackrel{X_i \text{ indep}}{=} \prod_{i=1}^n P(X = x_i) \\ &\stackrel{X_i \sim X}{=} \prod_{i=1}^n \left(\frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \right) \\ &= \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}, \lambda > 0 \end{aligned}$$

Passo 2 — Função de log-verosimilhança

$$\ln L(\lambda|\underline{x}) = -n\lambda + \ln(\lambda) \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!)$$

Passo 3 — Maximização

A estimativa de MV de λ é aqui representada por $\hat{\lambda}$ e

$$\hat{\lambda} : \begin{cases} \left. \frac{d \ln L(\lambda|\underline{x})}{d\lambda} \right|_{\lambda=\hat{\lambda}} = 0 & \text{(ponto de estacionaridade)} \\ \left. \frac{d^2 \ln L(\lambda|\underline{x})}{d\lambda^2} \right|_{\lambda=\hat{\lambda}} < 0 & \text{(ponto de máximo)} \\ \begin{cases} -n + \frac{\sum_{i=1}^n x_i}{\hat{\lambda}} = 0 \\ -\frac{\sum_{i=1}^n x_i}{\hat{\lambda}^2} < 0 \end{cases} \\ \begin{cases} \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i \\ \text{Proposição verdadeira desde que } \sum_{i=1}^n x_i > 0 \end{cases} \end{cases}$$

Passo 4 — Estimativa de MV de λ

$$\begin{aligned} \hat{\lambda} &= \frac{1}{n} \sum_{i=1}^n x_i \\ &= \frac{10}{85} \\ &= \frac{2}{17} \end{aligned}$$

• Outro parâmetro DESCONHECIDO

$$\begin{aligned} h(\lambda) &= P(X \geq 1) \\ &= 1 - P(X = 0) \\ &= 1 - e^{-\lambda} \end{aligned}$$

• Estimativa de MV de $h(\lambda)$

Invocando a propriedade de invariância dos estimadores de máxima verosimilhança, pode concluir-se que a estimativa de MV de $h(\lambda) = 1 - e^{-\lambda}$ é igual a

$$\begin{aligned} \widehat{h(\lambda)} &= h(\hat{\lambda}) \\ &= 1 - e^{-\hat{\lambda}} \\ &= 1 - e^{-\frac{2}{17}} \\ &\simeq 0.110990. \end{aligned}$$

Exercício 3

• V.a. de interesse

X = aumento de peso em porcos alimentados com a ração A durante 20 semanas

Y = aumento de peso em porcos alimentados com a ração B durante 20 semanas

• Situação

$$X \sim \text{Normal}(\mu_X, \sigma_X^2) \perp\!\!\!\perp Y \sim \text{Normal}(\mu_Y, \sigma_Y^2)$$

$(\mu_X - \mu_Y)$ DESCONHECIDO

σ_X^2 e σ_Y^2 desconhecidos, no entanto, assume-se que são IGUAIS: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$

$n_X \leq 30$ ou $n_Y \leq 30$

- **Hipóteses**

$$H_0 : \mu_X - \mu_Y = \mu_0 = 0$$

$$H_1 : \mu_X - \mu_Y > \mu_0$$

- **Nível de significância**

$$\alpha_0 = 10\%$$

- **Estatística de teste**

$$T = \frac{(\bar{X} - \bar{Y}) - \mu_0}{\sqrt{\frac{(n_X-1)S_X^2 + (n_Y-1)S_Y^2}{n_X+n_Y-2} \times \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \sim_{H_0} t_{(n_X+n_Y-2)}$$

dado que se pretende efectuar um teste sobre a diferença de valores esperados de duas populações normais independentes, com variâncias desconhecidas mas que se assume serem iguais.

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

Estamos a lidar com um teste unilateral superior ($H_1 : \mu_X - \mu_Y > \mu_0$), logo a região de rejeição de H_0 (para valores da estatística de teste) é um intervalo do tipo $W = (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 | H_0) = \alpha_0$, i.e.,

$$\begin{aligned} c &= F_{t_{(n_X+n_Y-2)}}^{-1}(1 - \alpha_0) \\ &= F_{t_{(6+7-2)}}^{-1}(1 - 0.10) \\ &= F_{t_{(11)}}^{-1}(0.9) \\ &\stackrel{\text{tabela}}{=} 1.363. \end{aligned}$$

- **Decisão**

Atendendo a que $\sum_{i=1}^7 x_i = 490$, $\sum_{i=1}^7 x_i^2 = 34\,432$, $\sum_{i=1}^6 y_i = 384$ e $\sum_{i=1}^6 y_i^2 = 24\,790$, segue-se:

$$\circ n_X = 6$$

$$\bar{x} = \frac{1}{n_X} \sum_{i=1}^{n_X} x_i = \frac{490}{7} = 70$$

$$s_X^2 = \frac{1}{n_X-1} \left[\left(\sum_{i=1}^{n_X} x_i^2 \right) - n_X (\bar{x})^2 \right] = \frac{1}{7-1} (34\,432 - 7 \times 70^2) = 22$$

$$n_Y = 6$$

$$\bar{y} = \frac{1}{n_Y} \sum_{i=1}^{n_Y} y_i = \frac{384}{6} = 64$$

$$s_Y^2 = \frac{1}{n_Y-1} \left[\left(\sum_{i=1}^{n_Y} y_i^2 \right) - n_Y (\bar{y})^2 \right] = \frac{1}{6-1} (24\,790 - 6 \times 64^2) = 42.8.$$

O valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \frac{(\bar{x} - \bar{y}) - \mu_0}{\sqrt{\frac{(n_X-1)s_X^2 + (n_Y-1)s_Y^2}{n_X+n_Y-2} \times \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \\ &= \frac{(70 - 64) - 0}{\sqrt{\frac{(7-1) \times 22 + (6-1) \times 42.8}{7+6-2} \left(\frac{1}{7} + \frac{1}{6}\right)}} \\ &\simeq 1.923. \end{aligned}$$

Como $t = 1.923 \in W = (1.363, +\infty)$, devemos rejeitar H_0 a qualquer n.s. $\alpha \geq 10\%$.

Exercício 1

(a) • **V.a. de interesse**

X = classificação de uma criança quanto ao género e à PHDA

• **Hipóteses**

Sejam

$$p_i = P(X = i)$$

$$= \begin{cases} \theta/2, & i = 1 \\ (1 - \theta)/2, & i = 2 \\ \theta(2 - \theta)/2, & i = 3 \\ (1 - \theta)^2/2, & i = 4 \end{cases}$$

$$p_i^0 \stackrel{\theta=0.9}{=} \begin{cases} 0.9/2, & i = 1 \\ (1 - 0.9)/2, & i = 2 \\ 0.9 \times (2 - 0.9)/2, & i = 3 \\ (1 - 0.9)^2/2, & i = 4. \end{cases}$$

Então confrontar-se-ão as seguintes hipóteses:

$$H_0 : p_i = p_i^0, i = 1, \dots, 4$$

$$H_1 : \exists i : p_i \neq p_i^0$$

• **Estatística de Teste**

$$T = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \underset{H_0}{\sim} \chi_{(k-\beta-1)}^2,$$

onde:

k = No. de classes = 4

O_i = Frequência absoluta observável da classe i

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

β = 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 $W = (c, +\infty)$.

• **Frequências absolutas esperadas sob H_0**

Atendendo a que, sob H_0 , $p_i = p_i^0$, as frequências absolutas esperadas sob H_0 são iguais a

$$E_i = n \times p_i^0$$

$$= \begin{cases} 2000 \times \frac{0.9}{2} = 900, & i = 1 \\ 2000 \times \frac{1-0.9}{2} = 100, & i = 2 \\ 2000 \times \frac{0.9 \times (2-0.9)}{2} = 990, & i = 3 \\ 2000 \times \frac{(1-0.9)^2}{2} = 10, & i = 4. \end{cases}$$

[Importa notar que não é necessário qualquer agrupamento de classes uma vez que em pelo menos 80% das classes se verifica $E_i \geq 5$.]

• **Decisão**

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

¹Dado que a distribuição conjecturada em H_0 está completamente especificada, i.e., H_0 é uma hipótese simples.

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	{1}	884	900	$\frac{(884-900)^2}{900} \simeq 0.284$
2	{2}	86	100	1.96
3	{3}	1018	990	0.792
4	{4}	12	10	0.4
		$\sum_{i=1}^k o_i = n = 2000$	$\sum_{i=1}^k E_i = n = 2000$	$t = \sum_{i=1}^k \frac{(o_i - E_i)^2}{E_i} \simeq 3.436$

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

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

$$\begin{aligned} p\text{-value} &= P(T > t \mid H_0) \\ &= P(T > 3.436 \mid H_0) \\ &\simeq 1 - F_{\chi^2_{(4-1-0)}}(3.436) \end{aligned}$$

Recorrendo às tabelas de quantis da distribuição do qui-quadrado podemos adiantar um intervalo para o $p\text{-value}$. Com efeito,

$$\begin{aligned} F_{\chi^2_{(3)}}^{-1}(0.600) = 2.946 &< 3.436 < 3.665 = F_{\chi^2_{(3)}}^{-1}(0.700) \\ 0.600 &< F_{\chi^2_{(3)}}(3.436) < 0.700 \\ 0.300 = 1 - 0.700 &> 1 - F_{\chi^2_{(3)}}(3.436) < 1 - 0.600 = 0.400. \end{aligned}$$

Deste modo, o intervalo aproximado para o $p\text{-value}$ é $(0.300, 0.400)$, pelo que

- não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 30\%$, nomeadamente aos níveis usuais de significância de 1%, 5% e 10%;
- devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 40\%$

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

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

$$\begin{aligned} p\text{-value} &= P(T > t \mid H_0) \\ &= P(T > 3.436 \mid H_0) \\ &\simeq 1 - F_{\chi^2_{(3)}}(3.436) \\ &\simeq 0.329158, \end{aligned}$$

pelo que

- não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 32.9158\%$, nomeadamente aos níveis usuais de significância de 1%, 5% e 10%;
- rejeitar H_0 a qualquer n.s. $\alpha_0 > 32.9158\%$.

Exercício 2

(a) • **[Modelo de RLS**

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

Y_i = resultado da medição da i – ésima concentração conhecida

x_i = valor da i – ésima concentração conhecida

ϵ_i = erro aleatório associado à medição da i – ésima concentração conhecida

• **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 β_0 e β_1**

Uma vez que

- $n = 5$
- $\sum_{i=1}^5 x_i = 50$
 $\bar{x} = 10$
 $\sum_{i=1}^n x_i^2 = 540$
 $\sum_{i=1}^n x_i^2 - n(\bar{x})^2 = 540 - 5 \times 10^2 = 40$
- $\sum_{i=1}^5 y_i = 150$
 $\bar{y} = 30$
 $\sum_{i=1}^n y_i^2 = 4676$
 $\sum_{i=1}^n y_i^2 - n(\bar{y})^2 = 4676 - 5 \times 30^2 = 176$
- $\sum_{i=1}^n x_i y_i = 1582$
 $\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} = 1582 - 5 \times 10 \times 30 = 82,$

as estimativas de β_1 e β_0 são, para este modelo, iguais a:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2} \\ &= \frac{82}{40} \\ &= 2.05 \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\ &= 30 - 2.05 \times 10 \\ &= 9.5\end{aligned}$$

• **Recta de regressão**

$$\hat{E}(Y|x) = \hat{\beta}_0 + \hat{\beta}_1 \times x = 9.5 + 2.05 \times x, \quad x \in [\min x_i, \max x_i]$$

(b) • **Passo 1 — Seleção da v.a. fulcral para β_1**

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

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

Ao ter em conta que, neste caso, $n = 5$ e $(1 - \alpha) \times 100\% = 95\%$, i.e., $\alpha = 0.05$, usaremos os quantis de probabilidade:

$$\begin{cases} a_\alpha = -F_{t_{(n-2)}}^{-1}(1 - \alpha/2) = -F_{t_{(3)}}^{-1}(0.975) \stackrel{\text{tabela}}{=} -3.182 \\ b_\alpha = F_{t_{(n-2)}}^{-1}(1 - \alpha/2) = F_{t_{(3)}}^{-1}(0.975) \stackrel{\text{tabela}}{=} 3.182 \end{cases}$$

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

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

...

$$P\left[\hat{\beta}_1 - F_{t_{(n-2)}}^{-1}(1 - \alpha/2) \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}} \hat{\beta}_1 + F_{t_{(n-2)}}^{-1}(1 - \alpha/2) \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}}\right] = 1 - \alpha.$$

• **Passo 4 — Concretização**

Atente-se que

- $\hat{\beta}_1 = 2.05$
- $F_{t_{(n-2)}}^{-1}(1 - \alpha/2) = 3.182$
- $\sum_{i=1}^n x_i^2 - n(\bar{x})^2 = 40.$

Se a isto acrescentarmos que

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-2} \left[\left(\sum_{i=1}^n y_i^2 - n\bar{y}^2 \right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \right] \\ &= \frac{1}{5-2} (176 - (2.05)^2 \times 40) \\ &\simeq 2.633 \\ IC_{(1-\alpha) \times 100\%}(\beta_1) &= \left[\hat{\beta}_1 \pm F_{t(n-2)}^{-1}(1-\alpha/2) \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}} \right]\end{aligned}$$

temos:

$$\begin{aligned}IC_{95\%}(\beta_1) &\simeq \left[2.05 \pm 3.182 \times \sqrt{\frac{2.633}{40}} \right] \\ &\simeq [2.05 \pm 0.7955] \\ &= [1.2545, 2.8455].\end{aligned}$$

(c) • **Significância da regressão**

Pretende testar-se $H_0 : \beta_1 = \beta_{1,0} = 0$ (regressão não é significativa, i.e., a variável resposta Y não depende da variável explicativa x) contra $H_0 : \beta_1 = \beta_{1,0} \neq 0$ (regressão é significativa).

Invocando a relação entre intervalos de confiança e testes de hipóteses e dado que

$$\beta_{1,0} = 0 \notin IC_{95\%}(\beta_1) = [1.2545, 2.8455],$$

a hipótese H_0 deve ser rejeitada ao nível de significância $\alpha = 100\% - 95\% = 5\%$.²

²Mais, devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 5\%$.