Probabilidades e Estatística

LEAN, LEGI, LEGM, MEAmbi, MEC

1º semestre – 2019/2020
09/01/2020 – **11:00**

Duração: 90 minutos **2º Teste B**

Justifique convenientemente todas as respostas

Grupo I 10 valores

1. Suponha que a duração de fusíveis (em ano) é modelada pela variável aleatória X com função de densidade de probabilidade

$$f_X(x) = \begin{cases} \frac{\lambda^3}{2} x^2 e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

onde λ ($\lambda > 0$) é um parâmetro desconhecido. Seja $(X_1, \ldots, X_n)$ uma amostra aleatória de X.

(a) Deduza o estimador de máxima verosimilhança de λ, com base na amostra aleatória acima. (3.0)

- **V.a. de interesse**
  X = duração (em anos) de fusíveis
- **F.d.p. de** X
  $$f(x) = \begin{cases} \frac{\lambda^3}{2} x^2 e^{-\lambda x}, & x \geq 0, \\ 0, & \text{caso contrário} \end{cases}$$
- **Parâmetro desconhecido**
  $\lambda, \quad \lambda > 0$
- **Amostra**
  $\underline{x} = (x_1, \ldots, x_n)$ amostra de dimensão n proveniente da população X.
- **Obtenção do estimador de MV de** λ

  Passo 1 — Função de verosimilhança

$$\begin{aligned} L(\lambda \mid \underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\overset{X_i\, indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \\ &\overset{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) \\ &= \prod_{i=1}^{n} \left[\frac{\lambda^3}{2} x_i^2 e^{-\lambda x_i} \right] \\ &= \frac{\lambda^{3n}}{2^n} \left(\prod_{i=1}^{n} x_i \right)^2 e^{-\lambda \sum_{i=1}^{n} x_i}, \quad \lambda > 0 \end{aligned}$$

  Passo 2 — Função de log-verosimilhança

$$\ln L(\lambda \mid \underline{x}) = 3n \ln(\lambda) - n \ln(2) + 2 \sum_{i=1}^{n} \ln(x_i) - \lambda \sum_{i=1}^{n} x_i$$

  Passo 3 — Maximização
  A estimativa de MV de λ é doravante representada por $\hat{\lambda}$ e

$$\hat{\lambda} \; : \; \begin{cases} \left.\frac{d \ln L(\lambda|\underline{x})}{d\lambda}\right|_{\lambda=\hat{\lambda}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2 \ln L(\lambda|\underline{x})}{d\lambda^2}\right|_{\lambda=\hat{\lambda}} < 0 & \text{(ponto de máximo)} \end{cases}$$

$$\begin{cases} \frac{3n}{\hat{\lambda}} - \sum_{i=1}^{n} x_i = 0 \\ -\frac{3n}{\hat{\lambda}^2} < 0 \end{cases}$$

$$\hat{\lambda} \quad : \quad \begin{cases} \hat{\lambda} = \frac{3n}{\sum_{i=1}^n x_i} \\ -\frac{(\sum_{i=1}^n x_i)^2}{3n} < 0 \qquad \text{(proposição verdadeira [já que } \sum_{i=1}^n x_i > 0\text{]).} \end{cases}$$

Passo 4 — Estimador de MV de λ

$$EMV(\lambda) = \frac{3n}{\sum_{i=1}^n X_i} \qquad \left[\ = \frac{3}{\bar{X}} \ \right].$$

(b) Com base na amostra $(x_1, x_2, \ldots, x_{10})$, referente às durações de 10 fusíveis e que conduziu a $\sum_{i=1}^{10} x_i = 60$, obtenha a estimativa de máxima verosimilhança de $V(X) = \frac{3}{\lambda^2}$. (2.0)

- **Estimativa de MV de λ**

$$\begin{aligned} \hat{\lambda} &= \frac{3n}{\sum_{i=1}^n x_i} \qquad [= 3/\bar{x}] \\ &= \frac{3 \times 10}{60} \\ &= 0.5 \end{aligned}$$

- **Outro parâmetro desconhecido**

$$h(\lambda) = \frac{3}{\lambda^2}$$

- **Estimativa de MV de $h(\lambda)$**
  Ao invocar a propriedade de invariância dos estimadores de MV, concluímos que a estimativa de MV de $h(\lambda)$ é dada por

$$\begin{aligned} \widehat{h(\lambda)} &= h(\hat{\lambda}) \\ &= \frac{3}{\hat{\lambda}^2} \\ &= \frac{3}{0.5^2} \\ &= 12. \end{aligned}$$

2. Considere que a variável aleatória X representa o tempo de processamento de uma tarefa (em minuto) por certo tipo de máquina. Com base numa amostra casual de tempos de processamento, obteve-se $\sum_{i=1}^{10} x_i = 29.8$ e $\sum_{i=1}^{10} x_i^2 = 89.94$. Admita que X possui distribuição normal.

(a) Determine um intervalo de confiança a 99% para o desvio padrão da variável aleatória X. (2.0)

- **V.a. de interesse**
  X = tempo de processamento (em minutos) da tarefa
- **Situação**
  $X \sim \text{normal}(\mu, \sigma^2)$
  μ desconhecido
  σ^2 DESCONHECIDO
- **Obtenção de IC para σ**
  Passo 1 — Selecção da v.a. fulcral para σ

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

[dado que é suposto determinar um IC para o desvio padrão de uma população normal, com valor esperado desconhecido.]

Passo 2 — Obtenção dos quantis de probabilidade
Dado que $n = 10$ e $(1-\alpha) \times 100\% = 99\%$, usaremos os quantis

$$\begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(9)}}(0.005) \stackrel{tabela/calc.}{=} 1.735 \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(9)}}(0.995) \stackrel{tabela/calc.}{=} 23.59. \end{cases}$$

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 à expressão geral do IC para σ,

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

ao par de quantis acima e ao facto de

$$\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \frac{29.8}{10} \\ &= 2.98 \end{aligned}$$

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\,(\bar{x})^2\right] \\ &= \frac{1}{10-1}\left(89.94 - 10 \times 2.98^2\right) \\ &\simeq 0.126222, \end{aligned}$$

temos:

$$\begin{aligned} IC_{99\%}(\sigma) &\simeq \left[\sqrt{\frac{(10-1)\times 0.126222}{23.59}},\ \sqrt{\frac{(10-1)\times 0.126222}{1.735}}\right] \\ &\simeq [0.219445,\ 0.809169]. \end{aligned}$$

(b) Confronte as hipóteses $H_0: E(X) = 2.7$ e $H_1: E(X) > 2.7$. Decida com base no valor-p. (3.0)

- **V.a. de interesse e situação**
  Ver alínea a).

- **Hipóteses**
  $H_0: \mu = \mu_0 = 2.7$
  $H_1: \mu > \mu_0 = 2.7$

- **Estatística de teste**

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

  [dado que pretendemos efectuar um teste para o valor esperado de população com distribuição normal com variância desconhecida.]

- **Região de rejeição de H_0** (para valores de T)
  Estamos a lidar com um teste unilateral superior ($H_1: \mu > \mu_0$), logo a região de rejeição de H_0 é do tipo $W = (c, +\infty)$.

- **Decisão (com base num intervalo para o valor-p)**
  O valor observado da estatística de teste é igual a
$$\begin{aligned} t &= \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \\ &\simeq \frac{2.98-2.7}{\sqrt{\frac{0.126222}{10}}} \\ &\simeq 2.492240 \end{aligned}$$
  e a região de rejeição deste teste é um intervalo à direita. Logo temos
$$\begin{aligned} valor-p &= P(T>t \mid H_0) \\ &\simeq 1-F_{t_{(n-1)}}(t) \\ &\simeq 1-F_{t_{(9)}}(2.492240). \end{aligned}$$
  Ao recorrermos às tabelas de quantis da distribuição de $t_{(n-1)}$ podemos adiantar um intervalo para o valor-p. Para tal, basta enquadrar convenientemente $t=2.492240$ por dois quantis dessa mesma distribuição:
$$\begin{aligned} F^{-1}_{t_{(9)}}(0.975)=2.262 &< 2.492240 < 2.821=F^{-1}_{t_{(9)}}(0.99) \\ 1-0.99 &< 1-F_{t_{(9)}}(2.492240) < 1-0.975 \\ 0.01 &< valor-p < 0.025. \end{aligned}$$
  Por consequência:
  - não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 1\%$;
  - devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 2.5\%$, nomeadamente aos n.u.s. de 5% e 10%.

  [Ao dispormos de uma máquina de calcular gráfica, obtemos
$$valor-p = P(T>t \mid H_0) \simeq 1-F_{t_{(n-1)}}(t) \simeq 1-F_{t_{(9)}}(2.492240) \overset{calc.}{\simeq} 0.017148.$$
  Logo é suposto:
  - não rejeitar H_0 a qualquer nível de significância $\alpha_0 \leq 1.7148\%$, nomeadamente ao n.u.s. de 1%;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > 1.7148\%$, designadamente aos n.u.s de 5% e 10%.]

Grupo II	10 valores

1. Um engenheiro de produção defende a hipótese H_0 de que a voltagem de saída das baterias dos computadores é modelada por uma distribuição exponencial de parâmetro 0.2. A medição da voltagem de saída de 100 destas baterias, selecionadas aleatoriamente, conduziu à seguinte tabela de frequências:

Voltagem de saída	$[0,5]$	$]5,10]$	$]10,14]$	$]14,+\infty[$
Frequência absoluta observada	64	23	8	5
Frequência absoluta esperada sob H_0	E_1	23.254	7.453	E_4

(a) Calcule os valores das frequências absolutas esperadas E_1 e E_4 (aproximando-os às milésimas). (1.0)

- **V.a. de interesse**
  X = voltagem de saída de bateria de computadores

- **F.d. conjecturada**
  $F(x)=P(X\leq x)=1-e^{-0.2\,x}$, $x>0$.

- **Frequências absolutas esperadas omissas**
  Atendendo à dimensão da amostra $n=100$ e à f.d. conjecturada, temos:
$$\begin{aligned} E_1 &= n\times P(X\leq 5) \\ &= 100\times\left(1-e^{-0.2\times 5}\right) \\ &\simeq 63.212; \end{aligned}$$
$$E_4 = n\times P(X>14)$$

$$
\begin{aligned}
&= n - \sum_{i=1}^{3} e_i \\
&\simeq 100 - (63.212 + 23.254 + 7.453) \\
&= 6.081.
\end{aligned}
$$

(b) Teste a hipótese conjeturada pelo engenheiro, ao nível de significância de 5%. (3.0)

- **Hipóteses**
  $H_0 : X \sim \text{exponencial}(0.2)$
  $H_1 : X \not\sim \text{exponencial}(0.2)$

- **Nível de significância**
  $\alpha_0 = 5\%$

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

  $\beta =$ No. de parâmetros a estimar $= 0$.

- **Frequências absolutas esperadas sob H_0**
  De acordo com a tabela facultada e a alínea (a), as frequências absolutas esperadas sob H_0 aproximadas às centésimas são:

  $$E_1 \simeq 63.212; \quad E_2 \simeq 23.254; \quad E_3 \simeq 7.453; \quad E_4 \simeq 6.081.$$

  [Não é necessário fazer qualquer agrupamento de classes uma vez que se verifica $E_i \geq 5$, em pelo menos 80% das classes, e que $E_i \geq 1$, para todo o i. Caso fosse preciso efectuar agrupamento de classes, os valores de k e $c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0)$ teriam que ser recalculados...]

- **Região de rejeição de H_0** (para valores de T)
  Lidamos com um teste de ajustamento, logo a região de rejeição de H_0 é o intervalo à direita $W = (c, +\infty)$, onde
  $$
  \begin{aligned}
  c &= F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) \\
  &= F^{-1}_{\chi^2_{(4-0-1)}}(1-0.05) \\
  &\overset{tabela/calc.}{=} 7.815.
  \end{aligned}
  $$

- **Decisão**
  No cálculo do valor observado da estatística de teste convém recorrer à seguinte tabela auxiliar:

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esp. sob H_0 E_i	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$]0,5]$	64	63.212	$\frac{(64-63.212)^2}{63.212} \simeq 0.010$
2	$]5,10]$	23	23.254	$\frac{(23-23.254)^2}{23.254} \simeq 0.003$
3	$]10,14]$	8	7.453	0.040
4	$]14+\infty[$	5	6.081	0.192
		$\sum_{i=1}^{k} o_i = n = 100$	$\sum_{i=1}^{k} E_i = n = 100$	$t = \sum_{i=1}^{k} \frac{(o_i-E_i)^2}{E_i} \simeq 0.245$

  Uma vez que $t \simeq 0.245 \not\in W = (7.815, +\infty)$, não devemos rejeitar H_0 ao n.s. de $\alpha_0 = 5\%$ [nem a qualquer outro n.s. inferior a α_0].

2. O tempo de reparação (Y, em hora) de determinado equipamento depende do número de falhas sinalizadas remotamente (x). Dispõem-se dos seguintes valores, que se reportam a 12 observações de avarias independentes:

$$\bar{x} = 1.416667\,, \quad \sum_{i=1}^{12} x_i^2 = 59\,, \quad \bar{y} = 8.908333\,, \quad \sum_{i=1}^{12} y_i^2 = 2076.63\,, \quad \sum_{i=1}^{12} x_i y_i = 346.1,$$

onde $\left[\min_{i=1,\ldots,12} x_i,\, \max_{i=1,\ldots,12} x_i\right] = [1\,,6]$.

(a) Estime a reta de regressão linear simples dos mínimos quadrados. (2.0)

- **[Hipóteses de trabalho**
  $E(\epsilon_i) = 0,\ V(\epsilon_i) = \sigma^2,\ i = 1,\ldots,n$
  $cov(\epsilon_i, \epsilon_j) = 0,\ i \neq j,\ i,j = 1,\ldots,n]$

- **Estimativas de MQ de β_0 e β_1**
  Dado que

  $n = 12$

  $\bar{x} = 1.416667$
  $\sum_{i=1}^n x_i^2 = 59$
  $\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2 = 59 - 12 \times 1.416667^2 = 34.916655$

  $\bar{y} = 8.908333$
  $\sum_{i=1}^n y_i^2 = 2076.63$
  $\sum_{i=1}^n y_i^2 - n\,(\bar{y})^2 = 2076.63 - 12 \times 8.908333^2 = 1124.329238$

  $\sum_{i=1}^n x_i y_i = 346.1$
  $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 346.1 - 12 \times 1.416667 \times 8.908333 = 194.658303,$

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

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2} \\
&\simeq \frac{194.658303}{34.916655} \\
&\simeq 5.574941 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&\simeq 8.908333 - 5.574941 \times 1.416667 \\
&\simeq 1.010498.
\end{aligned}$$

  Assim a recta de regressão de mínimos quadrados é dada por:

$$\hat{\beta}_0 + \hat{\beta}_1 \times x = 1.010498 + 5.574941 \times x, \quad x \in \left[\min_{i=1,\ldots,12} x_i,\ \max_{i=1,\ldots,12} x_i\right] = [1\,,6]$$

(b) Admitindo a validade das hipóteses de trabalho habituais para o modelo de regressão linear simples de Y em x, obtenha um intervalo de confiança a 95% para o tempo esperado de reparação quando são sinalizadas 5 falhas. (3.0)

- **[Hipóteses de trabalho**
  $\epsilon_i \overset{i.i.d.}{\sim} \text{normal}(0, \sigma^2),\ i = 1,\ldots,n]$

- **Obtenção do IC para** $E(Y \mid x = x_0) = \beta_0 + \beta_1 x_0$, com $x_0 = 5$
  Passo 1 — V.a. fulcral para $E(Y \mid x = x_0) = \beta_0 + \beta_1 x_0$

$$Z = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \sim t_{(n-2)}$$

Passo 2 — Quantis de probabilidade
Já que $(1-\alpha)\times 100\% = 95\%$, temos $\alpha = 0.05$ e lidaremos com os quantis

$$\begin{cases} a_\alpha = F^{-1}_{t_{(n-2)}}(\alpha/2) = -F^{-1}_{t_{(12-2)}}(1-0.05/2) = -F^{-1}_{t_{(10)}}(0.975) \stackrel{tabela/calc.}{=} -2.228 \\ b_\alpha = F^{-1}_{t_{(12-2)}}(1-0.05/2) = F^{-1}_{t_{(10)}}(0.975) \stackrel{tabela/calc.}{=} 2.228. \end{cases}$$

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

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

$$P\left[a_\alpha \leq \frac{(\hat\beta_0+\hat\beta_1 x_0)-(\beta_0+\beta_1 x_0)}{\sqrt{\hat\sigma^2\times\left[\frac{1}{n}+\frac{(x_0-\bar x)^2}{\sum_{i=1}^n x_i^2-n\bar x^2}\right]}} \leq b_\alpha\right] = 1-\alpha$$

$$P\left[(\hat\beta_0+\hat\beta_1 x_0) - b_\alpha\times\sqrt{\hat\sigma^2\times\left[\frac{1}{n}+\frac{(x_0-\bar x)^2}{\sum_{i=1}^n x_i^2-n\bar x^2}\right]} \leq \beta_0+\beta_1 x_0 \right.$$
$$\left. \leq (\hat\beta_0+\hat\beta_1 x_0) - a_\alpha\times\sqrt{\hat\sigma^2\times\left[\frac{1}{n}+\frac{(x_0-\bar x)^2}{\sum_{i=1}^n x_i^2-n\bar x^2}\right]}\right] = 1-\alpha$$

- **Passo 4 — Concretização**
  Dado que a estimativa de σ^2 é igual a

$$\begin{aligned}\hat\sigma^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\bar y^2\right) - (\hat\beta_1)^2\left(\sum_{i=1}^n x_i^2 - n\bar x^2\right)\right] \\ &\simeq \frac{1}{12-2}\left(1124.329238 - 5.574941^2\times 34.916655\right) \\ &\simeq 3.912075\end{aligned}$$

  e a expressão geral do IC pretendido é

$$IC_{(1-\alpha)\times 100\%}(\beta_0+\beta_1 x_0) = \left[(\hat\beta_0+\hat\beta_1 x_0) \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\hat\sigma^2\times\left[\frac{1}{n}+\frac{(x_0-\bar x)^2}{\sum_{i=1}^n x_i^2-n\bar x^2}\right]}\right],$$

  temos

$$\begin{aligned}&IC_{95\%}(\beta_0+\beta_1\times 5) \\ &\simeq \left[(1.010498+5.574941\times 5)\pm 2.228\times\sqrt{3.912069\times\left[\frac{1}{12}+\frac{(5-1.416667)^2}{34.916655}\right]}\right] \\ &\simeq [28.885203 \pm 2.228\times 1.328396] \\ &\simeq [25.925537, 31.844869].\end{aligned}$$

(c) Calcule e interprete o coeficiente de determinação do modelo ajustado. (1.0)

- **Cálculo do coeficiente de determinação**

$$\begin{aligned} r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\bar x\bar y\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\bar x^2\right)\times\left(\sum_{i=1}^n y_i^2 - n\bar y^2\right)} \\ &= \frac{194.658303^2}{34.916655\times 1124.329238} \\ &\simeq 0.965205.\end{aligned}$$

- **Interpretação coeficiente de determinação**
  Cerca de 96.5% da variação total da variável resposta Y é explicada pela variável x, através do modelo de regressão linear simples ajustado, donde possamos afirmar que a recta estimada parece ajustar-se muito bem ao conjunto de dados.