DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA

Probabilidades e Estatística

LEBiol21, LEIC-A, LEQ21; LEIC-T

2º Semestre – 2022/2023
28/06/2023
13:00–15:00

Duração: **120** minutos — **Exame Época Normal – B**

- As questões de escolha múltipla têm uma e apenas uma resposta correta. Caso assinale uma resposta errada, serão descontados 0.5 valores.
- Se assinalar mais do que uma resposta numa questão de escolha múltipla, o resultado será classificado como errado e serão descontados 0.5 valores.
- As respostas às questões de escolha múltiplas têm de ser assinaladas na folha de instruções previamente distribuída.
- Para as restantes questões, terá de justificar convenientemente as suas respostas.

Pergunta 1 — 2 valores

O portal de reclamações de uma grande empresa recebe anualmente reclamações de 20% dos seus colaboradores, sendo que: entre os colaboradores que apresentam reclamações, 80% são do sexo feminino; entre os colaboradores que não apresentam reclamações, 70% são do sexo masculino.

Calcule a probabilidade de um colaborador da empresa, escolhido ao acaso, apresentar reclamação no período de um ano, sabendo que é do sexo feminino.

- **Quadro de acontecimentos e probabilidades**

  Para um colaborador da empresa escolhido ao acaso, temos

Acontecimento	Probabilidade
R = "colaborador apresenta reclamação no período de um ano"	$P(R) = 0.2$
F = "colaborador ser do sexo feminino"	$P(F) = ?$
$\bar{F}$ = "colaborador ser do sexo masculino"	
	$P(F \mid R) = 0.8$
	$P(\bar{F} \mid \bar{R}) = 0.7$

- **Cálculo da probabilidade pedida**

  Invocando o teorema de Bayes, temos

$$
\begin{aligned}
P(R \mid F) &= \frac{P(R) \times P(F \mid R)}{P(F)} = \frac{P(R) \times P(F \mid R)}{P(F \mid R) \times P(R) + P(F \mid \bar{R}) \times P(\bar{R})} \\
&= \frac{P(R) \times P(F \mid R)}{P(F \mid R) \times P(R) + [1 - P(\bar{F} \mid \bar{R})] \times [1 - P(R)]} = \frac{0.8 \times 0.2}{0.8 \times 0.2 + (1-0.7) \times (1-0.2)} = \frac{0.16}{0.4} = 0.4.
\end{aligned}
$$

Pergunta 2 — 2 valores

O tempo (em dias) entre avarias sucessivas de uma impressora utilizada num escritório é uma variável aleatória com distribuição exponencial com valor esperado igual a 15 dias.

Calcule a probabilidade de que decorram mais de 12 dias entre duas avarias consecutivas, sabendo que decorreram mais do que 10 dias entre estas avarias.

- **V.a. de interesse, f.d.p. e f.d.**

  X = tempo (em dias) entre avarias sucessivas de uma impressora $\qquad X \sim \text{exponencial}(\lambda)$

$$f_X(x) \overset{form.}{=} \begin{cases} \lambda\, e^{-\lambda\, x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt = \begin{cases} 0, & x < 0 \\ \int_0^x \lambda\, e^{-\lambda\, t}\, dt = 1 - e^{-\lambda\, x}, & x \geq 0 \end{cases}$$

- **Obtenção de λ**

  Tirando partido do formulário, temos $\quad \lambda \in \mathbb{R}+ : E(X) = 15 \Leftrightarrow \frac{1}{\lambda} = 15 \Leftrightarrow \lambda = \frac{1}{15}$.

- **Probabilidade pedida**

  Ao invocarmos a propriedade da falta de memória da distribuição exponencial, obtemos

$$\begin{aligned} P(X > 12 \mid X > 10) &= P(X > 10 + 2 \mid X > 10) \\ &= P(X > 2) \\ &= 1 - F_X(2) \\ &= e^{-2/15} \simeq 0.875173. \end{aligned}$$

  [Alternativamente,

$$P(X > 12 \mid X > 10) = \frac{P(X > 12, X > 10)}{P(X > 10)} = \frac{P(X > 12)}{P(X > 10)} = \frac{e^{-12\,\lambda}}{e^{-10\,\lambda}} = e^{-2\,\lambda} \equiv 1 - F_X(2) \simeq 0.875173.]$$

Pergunta 3	2 valores

Admita que a altitude de abertura automática de um pára-quedas de carga militar é uma variável aleatória X (em metros) com distribuição normal com valor esperado 175 e terceiro quartil 191.8625.

Qual é o desvio padrão de X? **A**: 24.0 **B**: 544.0 **C**: 25.0 **D**: 625.0

- **V.a.**

  X = altitude de abertura automática de um pára-quedas de carga militar

  $X \sim \text{normal}(\mu = 175, \sigma^2)$

- **Desvio padrão de X**

$$\sigma \;:\; F_X^{-1}(0.75) = 191.8625 \quad \Leftrightarrow \quad F_X(191.8625) = 0.75 \quad \Leftrightarrow \quad P\left(\frac{X-\mu}{\sigma} \leq \frac{191.8625-\mu}{\sigma}\right) = 0.75$$

$$\Phi\left(\frac{191.8625-\mu}{\sigma}\right) = 0.75 \quad \Leftrightarrow \quad \frac{191.8625-175}{\sigma} = \Phi^{-1}(0.75) \quad \Leftrightarrow \quad \sigma \overset{tabelas/calc.}{=} \frac{191.8625-175}{0.6745}$$

$$\sigma = 25.0.$$

Pergunta 4	2 valores

X_A e X_B representam a duração de vida (em minutos) do organismo A e do organismo B, respectivamente.

Considerando que X_A e X_B são variáveis aleatórias independentes, com distribuição normal, de valor esperado 120 e 80 (respectivamente), e desvio padrão 24 e 20 (respectivamente), determine a probabilidade da duração de vida do organismo A ser superior à do organismo B.

- **V.a.**

  X_i = duração de vida (em minutos) de microorganismo do tipo i, $i = A, B$

$$X_A \sim \text{normal}(E(X_A) = 120, V(X_A) = 24^2) \quad \perp\!\!\!\perp \quad X_B \sim \text{normal}(E(X_B) = 80, V(X_B) = 20^2)$$

- **Importante**

  $P(X_A > X_B) = P(X_A - X_B > 0)$, pelo que é conveniente lidar com a v.a. $(X_A - X_B)$, uma combinação linear de duas v.a. independentes com distribuição normal, logo também normalmente distribuída.

  $X_A - X_B \sim \text{normal}(E(X_A - X_B), V(X_A - X_B))$

  $E(X_A - X_B) = E(X_A) - E(X_B) = 120 - 80 = 40$

  $V(X_A - X_B) \overset{X_A \perp\!\!\!\perp X_B}{=} V(X_A) + V(X_B) = 24^2 + 20^2 = 976$

- **Probabilidade pedida**

$$\begin{aligned}
P(X_A > X_B) &= P(X_A - X_B > 0) \\
&= 1 - P(X_A - X_B \leq 0) \\
&= 1 - P\left[\frac{(X_A - X_B) - E(X_A - X_B)}{\sqrt{V(X_A - X_B)}} \leq \frac{0 - E(X_A - X_B)}{\sqrt{V(X_A - X_B)}}\right] \\
&= 1 - \Phi\left(\frac{0 - 40}{\sqrt{976}}\right) \\
&\simeq 1 - \Phi(-1.28) \\
&= \Phi(1.28) \\
&\overset{tabela/calc.}{=} 0.8997.
\end{aligned}$$

Pergunta 5	2 valores

Sejam X e Y duas variáveis aleatórias com função de probabilidade conjunta

$$P(X = x, Y = y) = \begin{cases} \frac{1}{12}, & \text{se } x, y = 0, 1, 2, 3 \text{ e } x \neq y \\ 0 & \text{caso contrário.} \end{cases}$$

Serão X e Y duas variáveis aleatórias independentes?

- **Par aleatório**

  (X, Y) com a f.p. conjunta do enunciado

- **Averiguação de (in)dependência**

  X e Y são v.a. INDEPENDENTES sse

$$P(X = y, Y = y) = P(X = x) \times P(Y = y), \quad \forall (x, y) \in \mathbb{R}^2.$$

  Uma vez que

$$\begin{aligned}
P(X = 0, Y = 0) &= 0 \\
&\neq \\
P(X = 0) \times P(Y = 0) &= [P(X = 0, Y = 1) + P(X = 0, Y = 2) + P(X = 0, Y = 3)] \\
&\quad \times [P(X = 1, Y = 0) + P(X = 2, Y = 0) + P(X = 3, Y = 0)] \\
&= (3 \times 1/12) \times (3 \times 1/12) = \frac{1}{4} \times \frac{1}{4} \\
&= \frac{1}{16},
\end{aligned}$$

  concluímos que X e Y não são v.a. INDEPENDENTES.

Pergunta 6	2 valores

Admita que X é uma variável aleatória com função de densidade de probabilidade

$$f_X(x) = \begin{cases} \frac{1+\theta x}{2}, & -1 \le x \le 1 \\ 0, & \text{outros valores de } x, \end{cases}$$

onde θ é um parâmetro desconhecido pertencente ao intervalo $[-1,1]$. A concretização de uma amostra aleatória de dimensão $n = 100$ proveniente de X conduziu à seguinte estimativa de máxima verosimilhança de θ, $\hat{\theta} = -0.0729$.

Qual é o valor da estimativa de máxima verosimilhança de $P(X < 0)$?

A: 0.46355 **B**: 0.53645 **C**: 0.481775 **D**: 0.518225

- **V.a. de interesse; f.d.p.**

  X

  $$f_X(x) = \begin{cases} \frac{1}{2}(1+\theta x), & -1 \le x \le 1 \\ 0, & \text{outros valores de } x \end{cases}$$

- **Parâmetro desconhecido**

  $\theta \quad (\theta \in [-1,1])$

- **Amostra**

  $\underline{x} = (x_1, \ldots, x_n)$, amostra de dimensão $n = 100$ proveniente da população X e tal que $\hat{\theta} = -0.0729$.

- **Outro parâmetro desconhecido**

  $$\begin{aligned} h(\theta) &= P(X < 0) \\ &= \int_{-\infty}^{0} f_X(x)\,dx \\ &= \int_{-1}^{0} \frac{1+\theta x}{2}\,dx \\ &= \left.\left(\frac{x}{2} + \frac{\theta x^2}{4}\right)\right|_{-1}^{0} \\ &= \frac{1}{2} - \frac{\theta}{4} \end{aligned}$$

- **Estimativa de MV de $h(\theta)$**

  Ao invocar a propriedade de invariância dos EMV, concluímos que a estimativa de MV de $h(\theta)$ é

  $$\begin{aligned} \widehat{h(\theta)} &= h(\hat{\theta}) \\ &= \frac{1}{2} - \frac{\hat{\theta}}{4} \\ &= \frac{1}{2} - \frac{-0.0729}{4} \\ &= 0.518225. \end{aligned}$$

Pergunta 7	2 valores

A massa (em kg) dos aparelhos de ondas de choque radiais portáteis é uma variável aleatória X com distribuição normal com valor esperado desconhecido e variância desconhecida. Foram selecionados aleatoriamente 21 desses aparelhos, tendo-se observado uma média e um desvio padrão amostrais de $\bar{x} = 10.5$ e $s = 1.3$ (respetivamente).

Determine um intervalo de confiança a 98% para o valor esperado de X.

- **V.a. de interesse**

  X = massa (em kg) de aparelho de ondas de choque radiais portátil

- **Situação**

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

  $\mu = E(X)$ DESCONHECIDO

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

- **Obtenção do IC para μ**

  Passo 1 — Variável aleatória fulcral para μ

$$Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}.$$

  Passo 2 — Quantis de probabilidade

  Dado que $n = 21$, $(1-\alpha) \times 100\% = 98\%$, usaremos os quantis

$$(a_\alpha, b_\alpha) : \begin{cases} P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2 \end{cases} \Leftrightarrow \begin{cases} a_\alpha = F^{-1}_{t_{(n-1)}}(\alpha/2) = F^{-1}_{t_{(20)}}(0.01) = -2.528 \\ b_\alpha = F^{-1}_{t_{(n-1)}}(1-\alpha/2) = F^{-1}_{t_{(20)}}(0.99) = 2.528. \end{cases}$$

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

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

$$P\left[a_\alpha \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq b_\alpha\right] = 1 - \alpha$$

$$P\left[\bar{X} - b_\alpha \times \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} - a_\alpha \times \frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

- **Passo 4 — Concretização**

  Atendendo à expressão geral do IC para μ,

$$IC_{(1-\alpha)\times 100\%}(\mu) \quad = \quad \left[\bar{x} \pm F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}}\right],$$

  ao par de quantis acima e ao facto de $n = 21$, $\bar{x} = 10.5$ e $s = 1.3$, temos:

$$\begin{aligned} IC_{98\%}(\mu) &= \left[10.5 - 2.528 \times \frac{1.3}{\sqrt{21}}, 10.5 + 2.528 \times \frac{1.3}{\sqrt{21}}\right] \\ &= [9.7828, 11.2172]. \end{aligned}$$

Pergunta 8	2 valores

Uma empresa conjetura que a probabilidade θ de um trabalhador escolhido ao acaso aceder a ligações em *e-mails* de remetentes desconhecidos é igual a 0.05. Para testar essa hipótese, foi realizada uma experiência em que se observou que 126 em 135 trabalhadores sorteados ao acaso não acederam a ligações em *e-mails* de remetentes desconhecidos.

Teste a hipótese $H_0 : \theta = 0.05$ contra $H_1 : \theta > 0.05$, calculando para o efeito o valor-p aproximado.

- **V.a. de interesse**

$$X = \begin{cases} 1, & \text{se trabalhador acede a ligações em } e\text{-}mails \text{ de remetentes desconhecidos} \\ 0, & c.c. \end{cases}$$

- **Situação**

  $X \sim \text{Bernoulli}(p)$

  $\theta = P(\text{trabalhador acede a ligações em } e\text{-}mails \text{ de remetentes desconhecidos})$ DESCONHECIDA

  $n = 135 \geq 30$ (suficientemente grande)

- **Hipóteses**

  $H_0 : \theta = \theta_0 = 0.05$ vs. $H_1 : \theta > \theta_0$

- **Estatística de teste**

$$T = \frac{\bar{X} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \overset{a}{\sim}_{H_0} \text{normal}(0,1).$$

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

  O teste é unilateral superior ($H_1 : \theta > \theta_0$), logo a região de rejeição de H_0 é do tipo $W = (c, +\infty)$.

- **Decisão (com base no valor-p)**

  Atendendo a que o valor observado da estatística de teste e o valor-p aproximado são iguais a

$$\begin{aligned} t &= \frac{\bar{x} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \\ &= \frac{\frac{135-126}{135} - 0.05}{\sqrt{\frac{0.05\,(1-0.05)}{135}}} \\ &\simeq 0.89 \\ valor - p &= P(T > t \mid H_0) \\ &\simeq 1 - \Phi(t) \\ &\simeq 1 - \Phi(0.89) \\ &\overset{tabelas/calc.}{=} 1 - 0.8133 \\ &= 0.1867. \end{aligned}$$

  Consequentemente, é suposto:

  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq valor - p = 18.67\%$, designadamente ao níveis de significância habituais ($\alpha_0 \in [0.01, 0.1]$, e.g., 1%, 5%, 10%)[;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > valor - p = 18.67\%$].

Pergunta 9 — 2 valores

Uma engenheira informática conjetura que o tamanho (MB) de ficheiros anexados em *e-mails* recebidos em determinado servidor possui função de distribuição dada por $F_0(x) = 1 - (x+1)^{-1/2}$, para $x > 0$ (hipótese H_0). A recolha de uma amostra casual de 47 *e-mails* conduziu à seguinte tabela de frequências:

Tamanho do ficheiro	$]0, 0.4]$	$]0.4, 2]$	$]2, 10]$	$]10, +\infty[$
Frequência absoluta observada	9	15	13	10
Frequência absoluta esperada sob H_0	7.28	12.59	12.96	14.17

Com base no teste de ajustamento do qui-quadrado, teste H_0 ao nível de significância de 5%.

- **V.a. de interesse**

  X = tamanho (MB) de ficheiros anexado

- **Hipóteses**

  $H_0 : F_X(x) = F_0(x) = F_0(x) = 1-(x+1)^{-1/2}$, $x > 0$

  $H_1 : F_X(x) \neq F_0(x) = F_0(x)$, para algum $x > 0$

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de teste**

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

  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.

- [**Frequências absolutas esperadas sob** H_0 — De acordo com a tabela facultada, as frequências absolutas esperadas sob H_0 aproximadas às centésimas são: $E_1 \simeq 7.28$; $E_2 \simeq 12.59$; $E_3 \simeq 12.96$; $E_4 \simeq 14.17$. 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 $c = F^{-1}_{\chi^2_{(k-1)}}(1-\alpha_0)$ teriam que ser recalculados...]

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

  Lidamos com um teste de ajustamento, donde a região de rejeição de H_0 é o intervalo à direita $W = (c, +\infty)$, onde

$$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(4-1)}}(1-0.05) \overset{tabela/calc.}{=} 7.815.$$

- **Decisão**

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	$]0,2]$	9	7.28	$\frac{(9-7.28)^2}{7.28} \simeq 0.406$
2	$]2,4]$	15	12.59	0.461
3	$]4,6]$	13	12.96	0.000
4	$]6,+\infty[$	10	14.17	1.227
		$\sum_{i=1}^k o_i = n = 47$	$\sum_{i=1}^k E_i = n = 47$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i} \simeq 2.094$

  Uma vez que $t \simeq 2.094 \not\in W = (7.815, +\infty)$, não devemos rejeitar H_0 ao n.s. de $\alpha_0 = 5\%$ [nem a qualquer outro n.s. inferior a α_0].

Pergunta 10 2 valores

Uma firma de material informático presta serviço de reparação de computadores para empresas. Os dados referentes ao número de computadores reparados (x) e ao tempo (Y, em minutos) despendido com as reparações em 18 empresas durante um trimestre conduziram aos resultados seguintes:

$$\sum_{i=1}^{18} x_i = 81, \quad \sum_{i=1}^{18} x_i^2 = 439, \quad \sum_{i=1}^{18} y_i = 1\,152, \quad \sum_{i=1}^{18} y_i^2 = 90\,232, \quad \sum_{i=1}^{18} x_i y_i = 6\,282, \quad \hat{\sigma}^2 \simeq 20.087248.$$

Admita que as variáveis x e Y estão relacionadas de acordo com o modelo de regressão linear simples, $Y = \beta_0 + \beta_1 x + \epsilon$.

Tendo presente as hipóteses de trabalho convenientes, teste a hipótese $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$, ao nível de significância de 5%.

- **[Modelo de RLS e hipóteses de trabalho**

  Y = tempo despendido com as manutenções/reparações (v.a. resposta)

  x = número de computadores mantidos/reparados (variável explicativa)

  $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$

  $\epsilon_i \stackrel{i.i.d.}{\sim}$ normal$(0, \sigma^2)$, $i = 1, \ldots, n$]

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0$

  $H_1 : \beta_1 \neq \beta_{1,0}$

- **Nível de significância**

  $\alpha_0 = 5\%$

- **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 de T)

  Estamos a lidar com um teste bilateral ($H_1 : \beta_1 \neq \beta_{1,0}$), logo a região de rejeição de H_0 é do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 \mid H_0) = \alpha_0$, i.e.,

$$c \quad = \quad F^{-1}_{t_{(n-2)}}(1 - \alpha_0/2) = F^{-1}_{t_{(18-2)}}(1 - 0.05/2) \stackrel{tabela/calc.}{\simeq} 2.120.$$

- **Decisão**

  Atendendo a que a estimativa de MQ de β_1 é dada por

$$\hat{\beta}_1 \quad = \quad \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{6\,282 - 18 \times \frac{81}{18} \times \frac{1\,152}{18}}{439 - 18 \times (\frac{81}{18})^2} = \frac{1\,098}{74.5} \simeq 14.738255$$

  e que $\hat{\sigma}^2 \stackrel{enunc.}{\simeq} 20.087248$, o valor observado da estatística de teste é igual a

$$t \quad = \quad \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \simeq \frac{14.73825 - 0}{\sqrt{\frac{20.087248}{74.5}}} \simeq 28.383362.$$

  Como $t \simeq 28.383362 \in W = (-\infty, -2.120) \cup (2.120, +\infty)$ devemos rejeitar H_0 ao n.s. de $\alpha_0 = 5\%$ [assim como a qualquer n.s. superior a 5%. Ou seja, há forte evidência a favor de H_1, i.e., o número de computadores mantidos/reparados influencia significativamente o tempo gasto com as manutenções/reparações].