DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA	**Probabilidade e Estatística**	2º Semestre – 2021/2022
	LEAN/LEM/LEAmb/LEGM LEIC-A LEIC-T LERC-LEE LEAer	06/07/2022
	LEBiol LEBiom LEEC LEMec LEQ	**13:00-15:00**

Duração: **120** minutos **Exame Época Normal - (b)**

Justifique convenientemente todas as respostas

Pergunta 1 2 valores

Um estudo relativo a uma população levou à conclusão de que a probabilidade de um indivíduo fumador dessa população desenvolver algum tipo de cancro ao longo da vida é de 17%.

Sabendo que a fracção de indivíduos fumadores na população é de 20% e que 12.2 % dos indivíduos da população desenvolvem algum tipo de cancro ao longo da vida, calcule a probabilidade de um indivíduo não fumador dessa população desenvolver algum tipo de cancro ao longo da vida.

- **Acontecimentos e probabilidades para um indivíduo escolhido ao acaso da população**

Acontecimento	**Probabilidade**
F = "indivíduo é fumador"	$P(F) = 0.2$
C = "indivíduo desenvolve algum tipo de cancro ao longo da vida"	$P(C) = 0.122$
	$P(C \mid F) = 0.17$
	$P(C \mid \overline{F}) = ?$

- **Cálculo da probabilidade pedida**

$$\begin{aligned}
& P(C) = P(F) \times P(C \mid F) + P(\overline{F}) \times P(C \mid \overline{F}) \qquad \text{(lei da probabilidade total)} \\
\Leftrightarrow\ & P(C) = P(F) \times P(C \mid F) + [1 - P(F)] \times P(C \mid \overline{F}) \\
\Leftrightarrow\ & 0.122 = 0.2 \times 0.17 + (1 - 0.2) \times P(C \mid \overline{F}) \\
\Leftrightarrow\ & P(C \mid \overline{F}) = \frac{0.122 - 0.2 \times 0.17}{1 - 0.2} \\
\Leftrightarrow\ & P(C \mid \overline{F}) = 0.11.
\end{aligned}$$

Pergunta 2 2 valores

Admita que a variável aleatória X representa o número de doentes graves que chegam, por dia, às urgências de um hospital da região de Lisboa e possui uma distribuição de Poisson. Assume-se que chega pelo menos um destes doentes por dia com probabilidade 0.95. Obtenha o valor médio de doentes graves que chegam por dia, e a probabilidade de chegarem mais de 3 doentes graves sabendo que chega pelo menos 1 destes.

- **V.a. de interesse**
  X = número de doentes graves que chegam, por dia, às urgências do hospital.

- **Distribuição**

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

- **Cálculo do parâmetro** $\lambda = E(X)$

$$\begin{aligned} 1 - P(X \geq 1) &= P(X = 0) = e^{-\lambda} \\ &= = 0.05, \end{aligned}$$

  pelo que $\lambda = -\log(0.05) \approx 3$.

- **Probabilidade pedida**

$$\begin{aligned} P(X > 3 | X \geq 1) &= \frac{1 - F(3)}{P(X \geq 1)} \\ &= \frac{1 - 0.6472}{0.95} \\ &= 0.3713. \end{aligned}$$

Pergunta 3	2 valores

Suponha que o tempo (em horas) que um engenheiro leva a realizar uma determinada tarefa é uma variável aleatória X que se distribui uniformemente entre 30 minutos e 2 horas. Dado que o engenheiro já trabalhou na tarefa pelo menos 1 hora, qual é a probabilidade de não demorar mais do que $\sqrt{E(1.5X^2)}$ horas para a concluir?

- **V.a. de interesse**
  X = Duração (em horas) da tarefa.

- **Distribuição**
  $X \sim \text{uniforme}(0.5, 2)$.

- **Probabilidade pedida**

$$\begin{aligned} E(X^2) &= V(X) + [E(X)]^2 \\ &= \frac{1.5^2}{12} + \frac{2.5^2}{2^2} \\ &= 1.75. \end{aligned}$$

$$\begin{aligned} P\left(X \leq \sqrt{E\left(1.5X^2\right)} | X \geq 1\right) &= P(X \leq 1.62019 | X \geq 1) \\ &= \frac{P(X \leq 1.62019 \wedge X \geq 1)}{P(X \geq 1)} \\ &= \frac{0.41346}{2/3} \\ &= 0.62019. \end{aligned}$$

Pergunta 4	2 valores

Seja (X, Y) um par aleatório com função densidade de probabilidade conjunta dada por

$$f_{X,Y}(x,y) = \begin{cases} 3x, & 0 < x < y < 1, \\ 0, & \text{outros casos.} \end{cases}$$

Determine a mediana de X quando $Y = \frac{1}{2}$.

- **Cálculo da distribuição marginal de** Y

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f_{X,Y}(x,y)\,dx \\ &= \int_0^y 3x\,dx \\ &= \frac{3y^2}{2}, 0 < y < 1, \end{aligned}$$

pelo que

$$f_Y(y) = \begin{cases} \frac{3y^2}{2}, & 0 < y < 1 \\ 0, & \text{outros casos} \end{cases}.$$

- **Cálculo da densidade condicionada de** $X|Y = 1/2$

$$\begin{aligned} f_{X|Y=1/2}(x) &= \frac{f_{X,Y}(x, 1/2)}{f_Y(1/2)} \\ &= 8x, 0 < x < 1/2, \end{aligned}$$

pelo que

$$f_{X|Y=1/2}(x) = \begin{cases} 8x, & 0 < x < 1/2 \\ 0, & \text{outros casos} \end{cases}.$$

- **Cálculo da distribuição condicionada de** $X|Y = 1/2$

$$\begin{aligned} F_{X|Y=1/2}(x) &= \int_{-\infty}^x f_{X|Y=1/2}(u)\,du \\ &= 4x^2, 0 < x < 1/2, \end{aligned}$$

pelo que

$$F_{X|Y=1/2}(x) = \begin{cases} 0 & x < 0, \\ 4x^2 & 0 < x < 1/2, \\ 1 & x > 1/2. \end{cases}$$

- **Mediana de** X **quando** $Y = \frac{1}{2}$

$$F_{X|Y=1/2}(x) = 1/2 \Leftrightarrow 4x^2 = 1/2 \wedge 0 < x < 1/2 \iff x = \frac{1}{\sqrt{8}}.$$

Pergunta 5	2 valores

Suponha que 10% dos residentes de uma cidade usam a bicicleta para fazerem as suas deslocações, de forma independente uns dos outros. Escreva a expressão que lhe permita calcular a probabilidade de pelo menos 3 dos 60 residentes inquiridos usem a bicicleta para fazer as suas deslocações. Calcule uma aproximação da probabilidade anterior, fazendo uso da distribuição normal.

- **V.a.** X_i

  Seja X_i uma variável aleatória que assume o valor 1 se o i-ésimo inquirido usa bicicleta nas suas deslocações e 0 caso contrário, $i = 1, \ldots, n$, com $n = 60$.

- **Distribuição** $X_i \sim \text{bernoulli}(p = 0.1)$, para $i = 1, \ldots, n$, são variáveis aleatórias iid.

- **Valor esperado de** X_i
  $\text{E}(X_i) = p = 0.1$.

- **Variância de** X_i
  $\text{Var}(X_i) = p(1-p) = 0.1 \times 0.9 = 0.09$.

- **V.a. de interesse**
  $S_n = \sum_{i=1}^n X_i$, número total de inquiridos que declarar usar bicicleta nas suas deslocações.

- **Distribuição exata de** S_n
  $S_n = \sum_{i=1}^n X_i \sim \text{binomial}(n = 60, p = 0.1)$.
  A expressão que permite calcular a probabilidade exata pedida é

$$P\left(\sum_{i=1}^{60} X_i \geq 3\right) = \sum_{i=3}^{60} \left(\begin{array}{c} n \\ i \end{array}\right) p^i (1-p)^{n-i}.$$

- **Valor esperado de** S_n $\text{E}(S_n) = \text{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{E}(X_i) = np = 60 \times 0.1 = 6$.

- **Variância de** S_n $\text{Var}(S_n) = \text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) = np(1-p) = 60 \times 0.1 \times 0.9 = 5.4$.

- **Distribuição aproximada de** S_n. De acordo com o teorema do limite central, temos

$$\frac{S_n - \text{E}(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - np}{\sqrt{np(1-p)}} \overset{a}{\sim} \text{normal}(0,1)$$

- **Probabilidade pedida (valor aproximado)**
  De acordo com o teorema do limite central, temos

$$\begin{aligned}
P\left(\sum_{i=1}^{60} X_i \geq 3\right) &= 1 - P\left(\sum_{i=1}^{60} X_i < 3\right) \\
&= 1 - P\left(\sum_{i=1}^{60} X_i \leq 2\right) \\
&= 1 - P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{2 - np}{\sqrt{np(1-p)}}\right) \\
&= 1 - P\left(\frac{S_n - 6}{\sqrt{5.4}} \leq \frac{2-6}{\sqrt{5.4}}\right) \\
&\overset{TLC}{\simeq} 1 - \Phi(-1.72) = \Phi(1.72) \\
&= 0.9573.
\end{aligned}$$

Pergunta 6	2 valores

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

$$f_X(x;\lambda) = \begin{cases} \frac{\lambda}{(1+x)^{\lambda+1}}, & x > 0, \\ 0, & \text{outros casos.} \end{cases}$$

com $\lambda > 0$ desconhecido. A concretização de uma amostra aleatória de dimensão 3 proveniente de X conduziu a $x_1 = 2$, $x_2 = 2.3$ e $x_3 = 0.8$. Obtenha a estimativa de máxima verosimilhança da $P(X > 1)$.

- Seja $\underline{x} = (x_1, x_2, x_3)$ uma amostra aleatória de dimensão 3 proveniente da população X.
- **Obtenção do estimador da máxima verosimilhança de λ**
  Passo 1 - Função de verosimilhança

$$\begin{aligned} L(\theta|\underline{x}) &= f_{\underline{X}}(\underline{x}) \overset{X_i\ indep}{=} \prod_{i=1}^{3} f_{X_i}(x_i) \overset{X_i \sim X}{=} \prod_{i=1}^{3} f_X(x_i) \\ &= \prod_{i=1}^{3} \lambda(1+x_i)^{-(\lambda+1)} = \lambda^3 \left(\prod_{i=1}^{3}(1+x_i)\right)^{-(\lambda+1)}. \end{aligned}$$

  Passo 2 - Função de log-verosimilhança

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

  Passo 3 - Maximização

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

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

$$\frac{d\ln L(\lambda|\underline{x})}{d\lambda} = 0 \Leftrightarrow \frac{3}{\lambda} - \sum_{i=1}^{3}\ln(1+x_i) = 0 \Rightarrow \hat{\lambda} = \frac{3}{\sum_{i=1}^{3}\ln(1+x_i)}.$$

$$\left.\frac{d^2\ln L(\lambda|\underline{x})}{d\lambda^2}\right|_{\lambda=\hat{\lambda}} = -\frac{3}{\hat{\lambda}^2} < 0, \text{ (proposição verdadeira)}$$

  Passo 4 - Estimador e estimativa de MV de λ

$$\text{Estimador } \hat{\lambda}_{MV} = \frac{3}{\sum_{i=1}^{3}\ln(1+X_i)}, \text{ Estimativa } \hat{\lambda} = \frac{3}{\sum_{i=1}^{3}\ln(1+x_i)} \approx 1.041.$$

- **Obtenção da estimativa de MV da $P(X > 1)$.**

$$P(X>1) = 1 - P(X<1) = 1 - \int_0^1 \lambda(1+x)^{-(\lambda+1)}dx = 2^{-\lambda}.$$

  Pela propriedade de invariância dos estimadores de máxima verosimilhança, concluímos que o estimativa de MV para $P(X > 1)$ é

$$\widehat{P(X>1)} = 2^{-1.041} \approx 0.4858.$$

Pergunta 7	2 valores

As medições de uma grandeza física obtidas com um dado instrumento distribuem-se de acordo com uma distribuição normal. Com base numa amostra de dimensão 61, em que se observou uma variância amostral de 1.3, determine um intervalo de confiança para o desvio padrão das medições a um nível de confiança de 90%.

- **V.a. de interesse**

  X = medição da grandeza física com um dado instrumento.

- **Situação** $X \sim \text{normal}(\mu, \sigma^2)$.

- **Seleção da variável aleatória fulcral para** σ^2

$$Z = \frac{60S^2}{\sigma^2} \sim \chi^2_{(60)}.$$

- **Obtenção dos quantis de probabilidade**

$$a_\alpha = F^{-1}_{\chi^2_{(60)}}(0.05) = 43.19;$$
$$b_\alpha = F^{-1}_{\chi^2_{(60)}}(0.95) = 79.08.$$

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

$$\begin{aligned} P(a_\alpha \leq Z \leq b_\alpha) &= 1-\alpha \\ &\Leftrightarrow \\ P\left(43.19 \leq \frac{60S^2}{\sigma^2} \leq 79.08\right) &= 0.90 \\ &\Leftrightarrow \\ P\left(\sqrt{\frac{60S^2}{79.08}} \leq \sigma \leq \sqrt{\frac{60S^2}{43.19}}\right) &= 0.90 \end{aligned}$$

- **Concretização**

$$\begin{aligned} IC_{90\%}(\sigma) &= \left[\sqrt{\frac{60s^2}{79.08}}, \sqrt{\frac{60s^2}{43.19}}\right] \\ &= \left[\sqrt{\frac{60 \times 1.3}{79.08}}, \sqrt{\frac{60 \times 1.3}{43.19}}\right] \\ &= [0.993, 1.344]. \end{aligned}$$

Pergunta 8	2 valores

Foram selecionadas 9 crianças de uma escola para um estudo sobre o quociente de inteligência de crianças com 7 anos de idade, tendo-se obtido uma média amostral de 107 e um desvio padrão amostral igual a 10. Assumindo que o quociente de inteligência segue uma distribuição normal, teste se o seu desvio padrão pode ser considerado como sendo igual 9, ou se pelo contrário é superior a 9, para um nível de significância de 5%.

- **V.a. de interesse**

  X= coeficiente de inteligência de crianças com 7 anos de idade.

- **Situação**

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

- **Hipóteses**

$$H_0 : \sigma = \sigma_0 = 9$$
$$H_1 : \sigma = \sigma_0 > 9$$

- **Estatística de teste**

$$T = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{(n-1)}.$$

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

  O teste é unilateral, logo a região de rejeição de H_0 é do tipo $W = (c, \infty)$.

- **Decisão (ao nível de significância de 5%)** Atendendo a que o valor observado da estatística de teste e o quantil de probabilidade $(100-5)\%$ são iguais a:

$$t = \frac{8 \times 10^2}{9^2} = 9.87654;$$
$$F^{-1}_{\chi_{(8)}}(0.95) = 15.51;$$

  pelo que não podemos rejeitar que desvio padrão é igual a 9, para um nível de significância de 5%.

Pergunta 9	2 valores

No sítio na Internet, uma unidade hoteleira da Serra da Estrela anuncia uma alta taxa de ocupação durante os fins de semana, pelo que aconselha a que as marcações sejam feitas com alguma antecedência. Em 100 tentativas independentes de marcação feitas em cima da hora, as frequências absolutas das recusas foram:

nº recusas	0	1	2	3	≥ 4
nº tentativas	13	14	7	4	62

Teste a hipótese do número de recusas por tentativa de marcação ter distribuição de Poisson com valor esperado igual a 2. Tome a decisão com base no valor-p do teste.

- **V.a. de interesse**

  X = Número de recusas por tentativa de marcação.

- **Hipóteses**

  $H_0 : X \sim \text{poisson}(2)$

  $H_1 : X \not\sim \text{poisson}(2)$

- **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 = 5

  O_i = Frequência absoluta observável da classe i

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

- **Cálculo do valor observado da estatística de teste**
  As frequências absolutas esperadas sob H_0 são dadas por $E_i = n \times p_i^0,\ i = 1, 2, \ldots, 5$, com

$$p_i^0 = P(X = i - 1 | H_0),$$

  e iguais a

$$E_1 = 100 \times P(X = 0 | H_0) = 100 \times 0.1353 = 13.53;$$
$$E_2 = 100 \times P(X = 1 | H_0) = 100 \times 0.2706 = 27.06;$$
$$E_3 = 100 \times P(X = 2 | H_0) = 100 \times 0.2706 = 27.06;$$
$$E_4 = 100 \times P(X = 3 | H_0) = 100 \times 0.1804 = 18.04;$$
$$E_5 = 100 \times (1 - (p_1^0 + p_2^0 + p_3^0 + p_4^0)) = 100 \times 0.1431 = 14.31.$$

$$\begin{aligned} t &= \sum_{i=1}^{5} \frac{(o_i - E_i)^2}{E_i} \\ &= \frac{(13 - 13.53)^2}{13.53} + \frac{(14 - 27.06)^2}{27.06} + \frac{(7 - 27.06)^2}{27.06} + \frac{(4 - 18.04)^2}{18.04} + \frac{(62 - 14.31)^2}{14.31} \\ &= 191.055. \end{aligned}$$

- **Decisão (com base no valor-p)** Uma vez que a região de rejeição de H_0 é para este teste um intervalo à direita temos:

$$\begin{aligned} valor - p &= P(T > q_0 \mid H_0) \\ &= P[T > 191.055 \mid H_0] \\ &\simeq 1 - F_{\chi^2_{(5-1)}}(191.055) \\ &\overset{calc.}{\simeq} 1 - 1 = 0 \end{aligned}$$

  Uma vez que valor-p = 0, os dados mostram evidência para rejeitar H_0 a qualquer nível de significância usual (1%, 5% e 10%).

Pergunta 10	2 valores

Considere o modelo de regressão linear simples, $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, com as hipóteses de trabalho usuais. Sabendo que para uma amostra de dimensão 16 a estimativa de mínimos quadrados de β_1 é igual a 0.35, $\sum_{i=1}^{16}(x_i - \bar{x})^2 = 100$ e $\sum_{i=1}^{16}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 32.2$, teste a hipótese $H_0 : \beta_1 = 0.2$ versus $H_1 : \beta_1 \neq 0.2$. Decida com base no valor-p.

- **Hipóteses de trabalho**

  No modelo de RLS, $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, consideraremos $\epsilon_i \overset{i.i.d.}{\sim} \text{Normal}(0, \sigma^2)$, $i = 1, \ldots, n$.

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0.2$

  $H_1 : \beta_1 \neq 0.2$

- **Estatística de teste**
  Assumindo que H_0 é verdadeira, pode escrever-se:

$$T_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim_{H_0} t_{(n-2)}.$$

- **Valor observado da estatística de teste**

  Tendo em conta que

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i^2 - \hat{y}_i\right)^2 \\
&= \frac{1}{n-2}\sum_{i=1}^{16}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \\
&= \frac{32.2}{16-2} \\
&= 2.3,
\end{aligned}$$

  o valor observado da estatística de teste é dado por

$$\begin{aligned}
t_0 &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}} \\
&= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \\
&= \frac{0.35 - 0.2}{\sqrt{\frac{2.3}{100}}} \\
&\simeq 0.989.
\end{aligned}$$

- **Cálculo do valor-p**

$$\begin{aligned}
p &= P(|T_0| \geq t_0 | H_0) \\
&= 1 - P(-0.99 < T_0 < 0.99 | H_0) \\
&= 1 - F_{t(14)}(0.99) + F_{t(14)}(-0.99) \\
&= 1 - F_{t(14)}(0.99) + 1 - F_{t(14)}(0.99) \\
&= 2 - 2F_{t(14)}(0.99) \\
&\simeq 2 - 2 \times 0.8305 = 0.3390,
\end{aligned}$$

Usando as tabelas apenas se consegue enquadrar o valor $F_{t(14)}(0.99)$:

$$\begin{array}{ccccc}
0.80 = F_{t(14)}(0.868) & < & F_{t(14)}(0.99) & < & F_{t(14)}(1.076) = 0.85 \\
2 \times 0.80 & < & 2F_{t(14)}(0.99) & < & 2 \times 0.85 \\
2 - 2 \times 0.85 & < & 2 - 2F_{t(14)}(0.99) & < & 2 - 2 \times 0.80 \\
0.3 & < & p & < & 0.4
\end{array}$$

- **Regra de decisão:**
  - Se $\alpha > 0.4 > p$, então rejeito H_0,
  - Se $\alpha < 0.3 < p$, então não rejeito H_0,
  - Caso contrário, não sei decidir.

- **Decisão**
  Como os níveis de significância usuais (0.01, 0.05 e 0.10) são inferiores 0.3 não rejeito H_0, i.e. não há evidências para questionar que o verdadeiro valor de β_1 é 0.2.