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	**10:30-12:30**

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

Justifique convenientemente todas as respostas

Pergunta 1 2 valores

Numa dada população, verificou-se que 20% dos indivíduos são atuais fumadores, 20% são antigos fumadores e os restantes são indivíduos que nunca foram fumadores. Para além disso, é sabido que o risco de um indivíduo da população desenvolver cancro do pulmão ao longo da vida é de 15% para atuais fumadores, 7% para antigos fumadores e 2% para indivíduos que nunca foram fumadores.

Retirado ao acaso um indivíduo da população, calcule a probabilidade de o mesmo desenvolver cancro do pulmão ao longo da vida.

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

Acontecimento	**Probabilidade**
$F =$ "o indivíduo é atual fumador"	$P(F) = 0.2$
$A =$ "o indivíduo é antigo fumador"	$P(A) = 0.2$
$N =$ "o indivíduo nunca foi fumador"	$P(N) = 1 - P(F) - P(A) = 0.6$
$D =$ "o indivíduo desenvolve cancro do pulmão ao longo da vida"	$P(D) = ?$
	$P(D \mid F) = 0.15$
	$P(D \mid A) = 0.07$
	$P(D \mid N) = 0.02$

- **Cálculo da probabilidade pedida**

$$\begin{aligned} P(D) &= P(F) \times P(D \mid F) + P(A) \times P(D \mid A) + P(N) \times P(D \mid N) \qquad \text{(lei da probabilidade total)} \\ &= 0.2 \times 0.15 + 0.2 \times 0.07 + 0.6 \times 0.02 \\ &= 0.056. \end{aligned}$$

Pergunta 2 2 valores

Admita que a variável aleatória X representa o número de doentes politraumatizados que chegam, por dia, às urgências de um hospital da região de Lisboa e que possui uma distribuição de Poisson. Considere que, em média, chegam a este serviço 4 doentes politraumatizados por dia.

(a) Qual é a probabilidade de chegarem mais de 10 doentes politraumatizados às urgências num dado dia?

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

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

  com $\lambda = 4$.
- **Probabilidade pedida**

  $$\begin{aligned} P(X > 10) &= 1 - F_X(10) \\ &= 1 - 0.9972 \\ &= 0.0028. \end{aligned}$$

(b) Considere 3 dias distintos em que o número de chegadas dos doentes politraumatizados àquele serviço são variáveis aleatórias independentes e identicamente distribuídas. Qual é a probabilidade de não chegar nenhum doente politraumatizado em pelo menos um destes 3 dias?

- **V.a. de interesse**
  Y = "número de dias, num total de 3, em que não chega nenhum doente"
  e seja
  $P(X = 0) = e^{-4}$.
- **Distribuição**

  $$Y \sim \text{binomial}(n = 3, p = e^{-4})$$

- **Probabilidade pedida**

  $$\begin{aligned} P(Y \geq 1) &= 1 - P(Y = 0) \\ &= 1 - (1 - e^{-4})^3 \\ &= 0.0539. \end{aligned}$$

Pergunta 3	2 valores

A procura semanal de gasolina num determinado posto, em dezenas de milhares de litros, é uma variável aleatória X com função densidade de probabilidade:

$$f(x) = \begin{cases} cx, & 0 < x < 1, \\ 1/4, & 1 \leq x < 4, \\ 0, & \text{outros casos.} \end{cases}$$

Mostre que, para $c = \frac{1}{2}$, essa função é, de facto, uma função densidade de probabilidade. Determine a $P(1.5 < X < E(X))$.

- **Cálculo da constante** c

$$\int_0^1 xc\,dx + \int_1^4 \frac{1}{4}\,dx = 1 \Leftrightarrow c = \frac{1}{2}.$$

- **Cálculo do** $E(X)$

$$\begin{aligned} E(X) &= \int_0^1 cx^2\,dx + \int_1^4 x\,\frac{1}{4}\,dx \\ &= \frac{49}{24}. \end{aligned}$$

- **Probabilidade pedida**

$$\begin{aligned} P\left(1.5 < X < \frac{49}{24}\right) &= \int_{1.5}^{\frac{49}{24}} 1/4\,dx \\ &= \frac{49}{96} - \frac{1.5}{4} \\ &= 0.1354. \end{aligned}$$

Pergunta 4	2 valores

O par aleatório discreto (X,Y), com suporte $\{(x,y) : x,y = 0,1,2\}$, apresenta as seguintes características: $P(X=0,Y=0) = P(X=1,Y=1) = 0$; $P(X=2) = P(X=0,Y=2) = P(X=1,Y=2) = P(X=2,Y=2) = 1/9$; a distribuição marginal de X é binomial $(2,p)$.

Determine o valor de p e a covariância entre X e Y. Serão X e Y independentes?

- **Distribuição conjunta do par aleatório** (X,Y)
Uma vez que se pretende calcular $cov(X,Y) = E(XY) - E(X)E(Y)$, serão necessários alguns cálculos auxiliares que envolverão as f.p. conjunta de (X,Y) e marginais de X e Y.
Como $X \sim$ binomial$(2,p)$ podemos determinar o valor de p usando
$$P(X=2) = \binom{2}{2} p^2 (1-p)^{2-2} = 1/9 \Longleftrightarrow p^2 = 1/9 \implies p = 1/3.$$
Facilmente se determinam as probabilidades marginais de X para $x=0$ e $x=1$. Assim como a f.p. marginal de Y. Então, a função de probabilidade conjunta do par (X,Y) pode ser representada por:

	Y			
X	0	1	2	$P(X=x)$
0	0	1/3	1/9	4/9
1	1/3	0	1/9	4/9
2	0	0	1/9	1/9
$P(Y=y)$	1/3	1/3	1/3	1

- **Valor esperado de XY**

$$\begin{aligned} E(XY) &= \sum_{x=0}^{2}\sum_{y=0}^{2} xy \times P(X=x, Y=y) \\ &= 1 \times 2 \times 1/9 + 2 \times 2 \times 1/9 \\ &= 2/3. \end{aligned}$$

- **Valor esperado de X**

$$E(X) = n \times p = 2 \times 1/3 = 2/3.$$

- **Valor esperado de Y**

$$\begin{aligned} E(Y) &= \sum_{y=0}^{2} y \times P(Y=y) \\ &= 0 \times 1/3 + 1 \times 1/3 + 2 \times 1/3 \\ &= 1. \end{aligned}$$

- **Covariância entre X e Y**

$$\begin{aligned} &= E(XY) - E(X)E(Y) \\ &= 2/3 - 2/3 \times 1 \\ &= 0. \end{aligned}$$

- **Conclusão**
  Visto que $cov(X,Y) = 0$ não podemos concluir nada sobre a independência de X e Y. Mas observa-se que $P(X=0, Y=0) \neq P(X=0) \times P(Y=0)$, logo $\exists_{(x,y)} : P(X=x, Y=y) \neq P(X=x) \times P(Y=y)$, logo as variáveis X e Y são dependentes.

Pergunta 5	2 valores

De acordo com o seu historial, um jogador de basquetebol tem probabilidade 0.8 de converter um lance livre. Se este jogador fizer 36 lances livres, de forma independente, indique um valor aproximado para a probabilidade de encestar pelo menos 27 dos lançamentos efetuados mas não mais de 31.

- **V.a.** X_i

  Seja X_i uma variável aleatória que assume o valor 1 se o jogador encesta e 0 se falha o i-ésimo lançamento efetuado, $i = 1, \ldots, n$, com $n = 36$.

- **Distribuição** $X_i \sim \text{bernoulli}(p = 0.8)$, para $i = 1, \ldots, n$, são variáveis aleatórias iid.

- **Valor esperado de** X_i
  $\mathrm{E}(X_i) = p = 0.8$.

- **Variância de** X_i
  $\mathrm{Var}(X_i) = p(1-p) = 0.8 \times 0.2 = 0.16$.

- **V.a. de interesse**
  $S_n = \sum_{i=1}^{n} X_i$, número total de lançamento encestados em 36 lançamentos.

- **Valor esperado de** S_n
  $\mathrm{E}(S_n) = \mathrm{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{E}(X_i) = np = 36 \times 0.8 = 28.8$.

- **Variância de** S_n
  $\mathrm{Var}(S_n) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = np(1-p) = 36 \times 0.8 \times 0.2 = 5.76$.

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

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

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

$$\begin{aligned}
P\left(27 \leq \sum_{i=1}^{36} X_i \leq 31\right) &= P\left(\frac{27-np}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{31-np}{\sqrt{np(1-p)}}\right) \\
&\overset{TLC}{\simeq} \Phi\left(\frac{31-28.8}{\sqrt{5.76}}\right) - \Phi\left(\frac{26-28.8}{\sqrt{5.76}}\right) \\
&= \Phi(0.9166667) - \Phi(-1.16) \\
&\simeq \Phi(0.92) - 1 + \Phi(1.16) \\
&= 0.8212 - 1 + 0.8770 \\
&= 0.6982.
\end{aligned}$$

Pergunta 6	2 valores

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

$$f_X(x;\mu) = \begin{cases} \frac{(\mu x)^{x-1}}{x!} e^{-\mu x}, & x = 1,2,3,\ldots \\ 0, & \text{outros casos} \end{cases}$$

com $\mu \in [0,1]$ desconhecido. A concretização de uma amostra aleatória de dimensão 3 proveniente de X conduziu a $x_1 = 1$, $x_2 = 4$ e $x_3 = 2$. Obtenha a estimativa de máxima verosimilhança do $E(X) = 1/(1-\mu)$.

- Seja $\underline{x} = (x_1, \ldots, x_n)$ uma amostra aleatória de dimensão n proviniente da população X.

- **Obtenção do estimador da máxima verosimilhança de μ**
Passo 1 - Função de verosimilhança

$$\begin{aligned} L(\mu|\underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\overset{X_i\ indep}{=} \prod_{i=1}^{3} f_{X_i}(x_i) \\ &\overset{X_i \sim X}{=} \prod_{i=1}^{3} f_{X_i}(x_i) \\ &= \prod_{i=1}^{3} \frac{(\mu x_i)^{x_i-1}}{x_i!} e^{-\mu x_i} \\ &= \mu^{\sum_{i=1}^{3} x_i - 3} \prod_{i=1}^{3} \frac{x_i^{x_i-1}}{x_i!} e^{-u\sum_{i=1}^{3} x_i}. \end{aligned}$$

Passo 2 - Função de log-verosimilhança

$$\ln L(\mu|\underline{x}) = \left(\sum_{i=1}^{3} x_i - 3\right)\ln(\mu) + \ln\left(\prod_{i=1}^{3} \frac{x_i^{x_i-1}}{x_i!}\right) - \mu \sum_{i=1}^{3} x_i.$$

Passo 3 - Maximização

A estimativa de MV de μ é doravante representada por $\hat{\mu}$ e

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

$$\frac{d \ln L(\mu|\underline{x})}{d\mu} = 0 \Leftrightarrow \frac{\sum_{i=1}^{3} x_i - 3}{\mu} = \sum_{i=1}^{3} x_i \Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{3} x_i - 3}{\sum_{i=1}^{n} x_i} = 1 - \frac{3}{\sum_{i=1}^{3} x_i} = 0.5714$$

$$\left.\frac{d^2 \ln L(\mu|\underline{x})}{d\mu^2}\right|_{\mu=\hat{\mu}} = -\frac{4}{0.5714^2} < 0, \text{ (proposição verdadeira)}$$

Pela propriedade de invariância dos estimadores de máxima verosimilhança, concluímos que a estimativa de MV para $E(X)$ é

$$\widehat{E(X)} = \frac{1}{1 - 0.5714} \approx 2.333.$$

Pergunta 7 — 2 valores

As medições de uma grandeza física obtidas com um dado instrumento distribuem-se de acordo com uma distribuição normal com desvio padrão igual a 1.4. Com base numa amostra aleatória de dimensão 64 em que se observou uma média amostral de 10.1, determine um intervalo de confiança para o valor esperado das medições a um nível de confiança de 95%.

- **V.a. de interesse**

  X = “medição da grandeza física com um dado instrumento"

- **Situação** $X \sim \text{normal}(\mu, \sigma^2 = 1.4^2)$.

- **Seleção da variável aleatória fulcral para** μ

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \text{normal}(0, 1).$$

- **Obtenção dos quantis de probabilidade**

$$a_\alpha = \Phi^{-1}(0.025) = -1.96;$$
$$b_\alpha = \Phi^{-1}(0.975) = 1.96.$$

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

$$\begin{array}{rcl} P(a_\alpha \leq Z \leq b_\alpha) & = & 1 - \alpha \\ & \Leftrightarrow & \\ P\left(-1.96 \leq \sqrt{n}\dfrac{\bar{X} - \mu}{\sigma} \leq 1.96\right) & = & 0.95. \end{array}$$

- **Concretização**

$$\begin{array}{rcl} IC_{95\%}(\mu) & = & \left[\bar{x} - 1.96\dfrac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\dfrac{\sigma}{\sqrt{n}}\right] \\ & = & \left[10.1 - 1.96\dfrac{1.4}{\sqrt{64}}, 10.1 + 1.96\dfrac{1.4}{\sqrt{64}}\right] \\ & = & [9.757, 10.443]. \end{array}$$

Pergunta 8	2 valores

Uma concentração de paracetamol superior a 150 mg por quilograma de massa corporal é considerada perigosa. Foram recolhidas 4 amostras a um paciente, tendo sido obtida uma concentração média amostral de 155 mg, e um desvio padrão amostral de 6 mg. Admitindo que a concentração de paracetamol por quilograma de massa corporal tem uma distribuição normal de desvio padrão igual a 5 mg, e que os valores registados no mesmo paciente podem ser considerados independentes, teste se podemos concluir que a concentração de paracetamol deste paciente pode ser considerada igual a 150 mg ou, se pelo contrário, este paciente está em risco devido a concentração perigosa, decidindo com base no valor-p.

- **V.a. de interesse**

  X= "concentração de paracetamol por quilograma de massa corporal"

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2 = 5^2)$.

- **Hipóteses**

  $$H_0 : \mu = \mu_0 = 150$$
  $$H_1 : \mu = \mu_0 > 150$$

- **Estatística de teste**

  $$T = \frac{\bar{X} - 150}{5/\sqrt{n}} \sim \text{normal}(0, 1).$$

- **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 (com base no valor-p)**

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

  $$t = \frac{155 - 150}{5/2} = 2$$
  $$\text{valor-p} = P(T > 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228,$$

  pelo que, para os níveis usuais de significância, rejeita-se H_0.

Pergunta 9	2 valores

Para estudar o fenómeno de desintegração radioativa do Bário-133 registaram-se, num período de 0.1 segundos, o número de partículas-α emitidas pelo Bário-133. Numa amostra casual de 60 períodos de 0.1 segundos, foram registadas as seguintes frequências:

Número de partículas-α	≤ 1	2	3	4	≥ 5
Frequência absoluta observada	17	14	17	7	5
Frequência absoluta esperada sob H_0	E_1	E_2	13.4	10.2	11.1

Teste a hipótese de o número de partículas emitidas no período referido ter distribuição Poisson de valor esperado 3, ao nível de significância de 5%. Comece por calcular as frequências absolutas esperadas sob H_0 omissas E_1 e E_2 (aproximando-as às décimas).

- **V.a. de interesse**

  X = "Número de partículas-α emitidas pelo Bário-133, num período de 0.1s"

- **Hipóteses**

  $H_0 : X \sim \text{poisson}(3)$

  $H_1 : X \not\sim \text{poisson}(3)$

- **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

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

  Ao efectuarmos um teste de ajustamento do qui-quadrado a região de rejeição de H_0 é um intervalo à direita $W = (c, +\infty)$, onde

$$c = F^{-1}_{\chi^2_{(k-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(5-1)}}(1-0.05) = F^{-1}_{\chi^2_{(4)}}(0.95) \overset{tabela/calc.}{=} 9.488.$$

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

  Para já, note-se que o conjunto de valores possíveis da distribuição Poisson(3) é $\mathbb{N}_0$. Assim, as classes a considerar são $\{0,1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5,6,\ldots\}$, como sugere a tabela de frequências do enunciado. Se, para além disso, atendermos a que f.d. de X, sob H_0, está tabelada, as frequências absolutas esperadas sob H_0 são iguais a:

$$
\begin{aligned}
E_i &= n \times p_i^0 \\
&= n \times P(X \in \text{ Classe } i \mid H_0) \\
&= n \times P[X \in \text{ Classe } i \mid X \sim \text{Poisson}(3)] \\
&= n \times \begin{cases} P[X \leq 1 \mid X \sim \text{Poisson}(3)], & i=1 \\ P[X = 2 \mid X \sim \text{Poisson}(3)], & i=2 \\ P[X = 3 \mid X \sim \text{Poisson}(3)], & i=2 \\ P[X = 4 \mid X \sim \text{Poisson}(3)], & i=2 \\ P[X \geq 5 \mid X \sim \text{Poisson}(3)], & i=3 \end{cases} \\
&= 60 \times \begin{cases} F_{Poisson(3)}(1) \overset{tabela}{=} 0.1991, & i=1 \\ F_{Poisson(3)}(2) - F_{Poisson(3)}(1) \overset{tabela}{=} 0.4232 - 0.1991 = 0.2241, & i=2 \\ 13.4, & i=3 \\ 10.2, & i=4 \\ 11.1, & i=5. \end{cases} \\
&= \begin{cases} 11.9, & i=1 \\ 13.4, & i=2 \\ 13.4, & i=3 \\ 10.2, & i=4 \\ 11.1, & i=5. \end{cases}
\end{aligned}
$$

- **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. 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	$\{0,1\}$	17	11.9	$\frac{(17-11.9)^2}{11.9} \simeq 2.1857$
2	$\{2\}$	14	13.4	$\frac{(14-13.4)^2}{13.4} \simeq 0.0268$
3	$\{3\}$	17	13.4	$\frac{(17-13.4)^2}{13.4} \simeq 0.9671$
3	$\{4\}$	7	10.2	$\frac{(7-10.2)^2}{10.2} \simeq 1.0039$
3	$\{5,6,\ldots\}$	5	11.1	$\frac{(5-11.1)^2}{11.1} \simeq 3.3522$
		$\sum_{i=1}^k o_i = n = 60$	$\sum_{i=1}^k E_i = n = 60$	$t = \sum_{i=1}^k \frac{(o_i - E_i)^2}{E_i} \simeq 7.5357$

Assim, temos

$$t_0 = \sum_{i=1}^{3} \frac{(o_i - E_i)^2}{E_i} \simeq 7.5357.$$

Uma vez que a $t_0 = 7.5357 \not\in W_{5\%} = (9.488, +\infty)$ não devemos rejeitar H_0 ao n.s. $\alpha_0 = 5\%$ [ou a qualquer n.s. inferior a 5%. Ou seja, a amostra não parece pôr em causa a premissa do número de partículas emitidas no período referido ter distribuição Poisson de valor esperado 3 para qualquer n.s. inferior ou igual a 5%.]

Pergunta 10	2 valores

Uma nutricionista está a investigar a relação entre o índice de colesterol total (x, em miligramas por decilitro) e o índice de massa corporal (Y, em quilogramas por altura ao quadrado). Numa amostra de 10 utentes de um centro de saúde, obtiveram-se os seguintes resultados:

$$\sum_{i=1}^{10} x_i = 2459,\quad \sum_{i=1}^{10} x_i^2 = 620155,\quad \sum_{i=1}^{10} y_i = 263.63,\quad \sum_{i=1}^{10} y_i^2 = 7091.388,\quad \sum_{i=1}^{10} x_i y_i = 65530.71,$$

onde $[\min_{i=1,\dots,10}(x_i), \max_{i=1,\dots,10}(x_i)] = [200, 310]$. Admita que x e Y estão relacionadas de acordo com o modelo de regressão linear simples, $Y = \beta_0 + \beta_1 x + \epsilon$. Obtenha a reta de mínimos quadrados com base nos dados fornecidos; além disso, calcule o coeficiente de determinação e interprete o seu valor.

Estimativas de MQ dos parâmetros desconhecidos β_0 e β_1

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{10} x_i y_i - 10 * \bar{x}\bar{y}}{\sum_{i=1}^{10} x_i^2 - 10 * \bar{x}^2} \\
&= \frac{65530.71 - 10 * 245.9 * 26.36}{620155 - 10 * 245.9^2} \\
&= \frac{704.093}{15486.9} \\
&= 0.0454.
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 * \bar{x} \\
&= 26.36 - 0.0454 * 245.9 \\
&= 15.196.
\end{aligned}$$

Coeficiente de determinação

$$\begin{aligned}
r^2 &= \frac{(\sum_{i=1}^{10} x_i y_i - 10 * \bar{x}\bar{y})^2}{(\sum_{i=1}^{10} x_i^2 - 10 * \bar{x}^2)(\sum_{i=1}^{10} y_i^2 - 10 * \bar{y}^2)} \\
&= \frac{704.093^2}{15486.9 * 141.3103} \\
&= 0.2265.
\end{aligned}$$

Interpretação do coeficiente de determinação

Cerca de 22% da variação total da variável resposta Y é explicada pela variável x, através do modelo de regressão linear simples ajustado.