Probabilidade e Estatística

1º Semestre – 2022/2023
08/02/2023
10:30-12:30

Duração: **120** minutos **Exame época recurso**
Versão A

- 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, será contado como resposta errada.
- As respostas às escolhas múltiplas são dadas na folha de instruções, previamente distribuída.
- Para as restantes questões, terá de justificar convenientemente as respostas.

Pergunta 1 2 valores

Num clube desportivo, praticam-se as modalidades de basquetebol e futebol, entre outras. Sabe-se que, considerando um atleta ao acaso deste clube, a probabilidade de ele praticar basquetebol é 1/5 e a probabilidade de ele praticar futebol é 2/5. Sabe-se ainda que, dos atletas que não praticam futebol, 3 em cada 4 não praticam basquetebol. Calcule a probabilidade de um qualquer atleta do clube praticar as duas modalidades desportivas.

- **Acontecimentos e probabilidades**

Acontecimento	Probabilidade
$A =$ "praticar basquetebol"	$P(A) = 0.2$
$B =$ "praticar futebol'	$P(B) = 0.4$
	$P(\overline{A}\vert\overline{B}) = 0.75$

- **Cálculo da probabilidade pedida**

$$\begin{aligned} P(\overline{A}|\overline{B}) &= \frac{1-P(A\cup B)}{1-P(B)} \Leftrightarrow P(A\cup B) = 1-P(\overline{B})\times P(\overline{A}|\overline{B}) = 0.55 \\ P(A\cap B) &= P(A)+P(B)-P(A\cup B) = 0.2+0.4-0.55 = 0.05 \end{aligned}$$

Pergunta 2 2.0 valores

Com a ingestão de um determinado fármaco para o tratamento da depressão, 20% dos doentes referem sentir efeitos secundários durante o tratamento. Calcule a probabilidade de ser necessário inquirir mais de 6 doentes até se encontrar um que sentiu efeitos secundários durante o tratamento sabendo que os 4 primeiros doentes já inquiridos não sentiram efeitos secundários durante o tratamento.

- **V.a. de interesse**

  X = Número de doentes inquiridos até encontrar um que sentiu efeitos secundários durante o tratamento

- **Distribuição**

$$X \sim Geo(0.2)$$

- **Probabilidade pretendida**

$$P(X > 6|X > 4) = P(X > 2) = 1 - P(X \leq 2) = 1 - [0.2 + 0.8 \times 0.2] = 0.64$$

  em que a 1ª igualdade decorre da propriedade da amnésia da distribuição geométrica.

Pergunta 3	2.0 valores

A quantidade de vinho (em dezenas de litros) que um produtor-engarrafador vende por dia é uma variável aleatória X com função de distribuição

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{50} & 0 \leq x < 5 \\ \frac{20x - x^2}{50} - 1 & 5 \leq x < 10 \\ 1 & x \geq 10 \end{cases}$$

A probabilidade do total de vendas não exceder 150% do valor esperado de X num dia em que já foram vendidos mais do que 4 dezenas de litros é igual a:

A: 0.816 B: 0.522 C: 0.516 D: 0.820

- **Função densidade de probabilidade**

$$f_X(x) = \begin{cases} \frac{x}{25} & 0 \leq x < 5 \\ \frac{2}{5} - \frac{x}{25} & 5 \leq x < 10 \\ 0 & \text{c.c.} \end{cases}$$

- **Valor esperado**

$$\mathbb{E}[X] = \int_0^5 x\frac{x}{25}dx + \int_5^{10} x\left(\frac{2}{5} - \frac{x}{25}\right)dx = 5$$

- **Probabilidade pedida**

$$P(X \leq 1.5 \times 5|X > 4) = \frac{F_X(7.5) - F_X(4)}{1 - F_X(4)} = 0.816$$

Pergunta 4	2.0 valores

Cada um de cinquenta números são arredondados (de forma independente) para o inteiro mais próximo e somados. Considere que os erros de arredondamento individuais são uniformemente distribuídos no intervalo (-0.5, 0.5). Obtenha um valor aproximado da probabilidade da soma resultante dos números arredondados exceder a soma dos valores exatos (não arredondados) mais do que 3 unidades.

- **V.a. de interesse**

$$X_i = \text{erro do } i\text{-ésimo arredondamento}$$

$$S = \text{erro da soma de 50 números} = \sum_{i=1}^{50} X_i$$

com $\{X_i, i = 1, \ldots, 50\}$ i.i.d. a $X \sim Unif(-0.5, 0.5)$.

- **Momentos**

$$E[X] = 0, Var[X] = \frac{1}{12}$$

pelo que

$$E[S] = 50 \times 0 = 0, Var[S] = 50 \times \frac{1}{12}$$

- **Probabilidade pedida** Pelo teorema do Limite Central:

$$P(S > 3) = 1 - P(S \le 3) = 1 - P\left(\frac{S}{\sqrt{\frac{50}{12}}} \le \frac{3}{\sqrt{\frac{50}{12}}}\right)$$
$$\approx 1 - \Phi(1.47) = 0.0708$$

Pergunta 5	2 valores

Seja (X, Y) uma variável aleatória que representa, para cada uma das famílias residentes numa determinada zona, o número de filhos da família (X) e o número de quartos do seu alojamento (Y). A função de probabilidade conjunta de (X, Y) encontra-se na tabela abaixo.

$Y \backslash X$	0	1	2
1	0.3	0.1	0.05
2	0.05	0.2	0.1
3	0.025	0.025	0.15

O número esperado de filhos por quarto é igual a:
A: 0.804 B 0.508 C: 0.925 D: 0.767

$$\mathbb{E}\left[\frac{X}{Y}\right] = \sum_{x=0}^{2}\sum_{y=1}^{3}\frac{x}{y}P(X = x, Y = y)$$
$$= 0.5083$$

Pergunta 6	2.0 valores

Seja X uma variável aleatória com função densidade de probabilidade

$$f_X(x) = (1+\theta)x^{\theta}, \quad x \in (0,1), \quad \theta > -1.$$

Determine o estimador de máxima verosimilhança de θ com base numa amostra de dimensão n.

- **Função de verosimilhança**

$$
\begin{aligned}
L(\theta; x_1, \ldots, x_n) &= \prod_{i=1}^{n} \left((1+\theta) x_i^{\theta}\right) \\
&= (1+\theta)^n \prod_{i=1}^{n} x_i^{\theta}
\end{aligned}
$$

- **Logaritmo e derivadas da função log- verosimilhança**

$$
\begin{aligned}
\log L(\theta; x_1, \ldots, x_n) &= n \ln(1+\theta) + \theta \sum_{i=1}^{n} \ln x_i \\
d(\log L(\theta; x_1, \ldots, x_n)) &= \frac{n}{1+\theta} + \sum_{i=1}^{n} \ln x_i \\
d^2(\log L(\theta; x_1, \ldots, x_n)) &= -\frac{n}{(1+\theta)^2} < 0, \quad \forall \theta
\end{aligned}
$$

pelo que o estimador de máxima verosimilhança de θ é o zero da primeira derivada, i.e.:

$$
EMV(\theta) = -\frac{n}{\sum_{i=1}^{n} \ln X_i} - 1
$$

Pergunta 7	2.0 valores

Admita que o tempo de execução de uma determinada tarefa numa obra segue uma distribuição normal, de valor esperado μ (desconhecido) e variância unitária. A partir de uma amostra de n trabalhadores que executaram esta tarefa obteve-se o seguinte intervalo de confiança a 95% para μ: [2.608; 3.392] minutos. Os valores de n e $\bar{x}$ correspondentes a esta amostra são, respectivamente:

A: $n = 36$ e $\bar{x} = 3.0$ B: $n = 25$ e $\bar{x} = 3.5$ C: $n = 36$ e $\bar{x} = 3.5$ D: $n = 25$ e $\bar{x} = 3.0$

- **V.a. de interesse**

$$
X \sim N(\mu, \sigma^2)
$$

com μ desconhecido.

- **Variável fulcral**

$$
\sqrt{n}\left(\bar{X} - \mu\right) \sim N(0,1)
$$

- **Quantis**

$$
\Phi^{-1}(0.025) = -1.96;\ \Phi^{-1}(0.975) = 1.96;
$$

- **ICA$_{0.95}(\mu)$**

$$
\left[\bar{X} - \frac{1.96}{\sqrt{n}}; \bar{X} + \frac{1.96}{\sqrt{n}}\right]
$$

- **Valores**

$$
\begin{cases} \bar{x} - \frac{1.96}{\sqrt{n}} = 2.608 \\ \bar{x} + \frac{1.96}{\sqrt{n}} = 3.392 \end{cases} = \begin{cases} \bar{x} = 3 \\ n = 25 \end{cases}
$$

Pergunta 8	2.0 valores

Com o intuito de investigar o tipo de audiência de certo programa televisivo seleccionaram-se de forma aleatória 25 espectadores, cuja idade foi registada, tendo-se obtido um desvio padrão amostral $s = 3.15$ anos.

Assuma que a idade de um espectador tem distribuição normal. Convencionando-se que a assistência é considerada heterogénea se a variância das idades ultrapassar os 6 anos, que conclui para um nível de significância de 5%?

- **V.a. de interesse e teste** $X \sim \text{Normal}(\mu, \sigma^2)$
  $\mathrm{H}_0 : \sigma^2 = 6$ e $\mathrm{H}_1 : \sigma^2 > 6$

- **Estatística de teste**

$$T_0 = \frac{24S^2}{6} \sim_a^{H_0} \chi^2_{(24)}$$

  O valor observado da estatística de teste é $t_0 = 39.69$

- **Região de rejeição e decisão**
  Rejeitar H_0 se valor observado da estatística de teste for superior a $F^{-1}_{\chi^2_{(24)}}(1-0.05) = 36.42$, pelo que a hipótese H_0 deverá ser rejeitada para $\alpha = 5\%$.

Pergunta 9	2.0 valores

De forma a decidir a cor da fachada de um monumento histórico de uma vila portuguesa, é feito um inquérito à população residente. Para o efeito foi recolhida uma amostra casual de 100 indivíduos dessa população. Os resultados abaixo indicam o número de vezes que cada cor proposta é escolhida como favorita.

Cor	Branco	Amarelo	Rosa	Azul
Nº votos	30	22	28	20

Teste a hipótese de equiprobabilidade da escolha das cores, ao nível de significância de 1%.

- **Variável aleatória de interesse**

$$X = \text{indicador da cor proposta} = \begin{cases} 1 & \text{Branco} \\ 2 & \text{Amarelo} \\ 3 & \text{Rosa} \\ 4 & \text{Azul} \end{cases}$$

- **Hipóteses**

$$\mathrm{H}_0 : P(X = i) = 0.25, \quad i \in \{1,2,3,4\}$$
$$\mathrm{H}_1 : \text{outra distribuição para } X$$

- **Estatística de teste**

$$T = \sum_{i=1}^{4} \frac{(E_i - O_i)^2}{E_i} \sim_a \chi^2_{(3)}$$

- **Valor observado da estatística de teste**

$$t_0 = \frac{(30-25)^2}{25} + \frac{(22-25)^2}{25} + \frac{(28-25)^2}{25} + \frac{(20-25)^2}{25} = 2.72$$

- **Região de rejeição** Rejeitar H_0 se o valor observado da estatística de teste, t_0, for superior a $F^{-1}_{\chi^2_{(3)}}(0.99) = 11.34$

- **Decisão** Não rejeitar a equiprobabilidade da escolha das cores, para $\alpha = 1\%$.

Pergunta 10	2.0 valores

É feita uma experiência em que se analisa o índice de octano da gasolina (Y) em função da quantidade de um novo aditivo (x). Para isso, foram realizados 6 ensaios. Assumindo um modelo de regressão linear simples, obtiveram-se os seguintes valores:

$$\sum_{i=1}^{6} x_i = 21, \quad \sum_{i=1}^{6} y_i = 496.8, \quad \sum_{i=1}^{6} x_i^2 = 91, \quad \sum_{i=1}^{6} y_i^2 = 41149.12, \quad \widehat{\beta}_1 = 0.886$$

O coeficiente de determinação é igual a:

A: 0.987 B: 0.867 C: 0.976 D: 0.901

Como $\widehat{\beta}_1 = \frac{\sum x_i Y_i - n\bar{x}\bar{Y}}{\sum x_i^2 - n\bar{x}^2}$, resulta que $\sum x_i Y_i = 0.886 \times \left(\sum x_i^2 - 6\bar{x}^2\right) + 6\bar{x}\bar{y} = 1754.3$, pelo que

$$r^2 = \frac{(1754.3 - 6 \times 21/6 \times 496.6/6)^2}{(91 - 6 \times (21/6)^2) \times (41149.12 - 6 \times (498.6)^2))} = 0.975$$