Probabilidade e Estatística

Todos os cursos

2º Semestre – 2021/2022
07/09/2022
10:30-12:30

Duração: **120** minutos **Exame Época Especial**

Justifique convenientemente todas as respostas

Pergunta 1	2 valores

Um vendedor de produtos informáticos instala em portáteis para venda processadores de tipos A, B e C nas proporções de 30%, 40% e 30%, respetivamente, oferecendo garantia de substituição de processadores por avaria no primeiro ano após a venda de portáteis. É conhecido que a probabilidade de uma processador avariar durante o primeiro ano de serviço é : 1% para processadores do tipo A; 1.5% para processadores do tipo B; e 2% para processadores do tipo C.

Calcule a probabilidade de um portátil, comprado ao vendedor, cujo processador instalado inicialmente no mesmo avaria durante o primeiro ano de serviço ter tido instalado inicialmente no mesmo um processador do tipo C.

- **Acontecimentos e probabilidades para uma portátil escolhida ao acaso**

Acontecimento	**Probabilidade**
A = "foi instalado inicialmente no portátil um processador do tipo A"	$P(A) = 0.3$
B = "foi instalado inicialmente no portátil um processador do tipo B"	$P(B) = 0.4$
C = "foi instalado inicialmente no portátil um processador do tipo C"	$P(B) = 0.3$
D = "o processador instalado inicialmente no portátil avaria durante o primeiro ano de serviço"	$P(D) =$?
	$P(D \mid A) = 0.01$
	$P(D \mid B) = 0.015$
	$P(D \mid C) = 0.02$

- **Cálculo da probabilidade pedida**

$$
\begin{aligned}
P(C \mid D) &= \frac{P(C) \times P(D \mid C)}{P(A) \times P(D \mid A) + P(B) \times P(D \mid B) + P(C) \times P(D \mid C)} \qquad \text{(teorema de Bayes)} \\
&= \frac{0.3 \times 0.02}{0.3 \times 0.01 + 0.4 \times 0.015 + 0.3 \times 0.02} \\
&= \frac{6}{15} = 0.4
\end{aligned}
$$

Pergunta 2	2 valores

Numa linha de montagem considere que o número de defeitos ocorridos findo o processo da montagem, é uma variável aleatória com função massa de probabilidade

$$
P(X = x) = \begin{cases} \frac{1}{10}, & x \in \{0, 1, 2, \ldots, 9\} \\ 0, & \text{outros casos} \end{cases}
$$

(a) Qual a probabilidade de se verificarem 3 defeitos quando é certo que pelo menos 3 irão ocorrer?

- **V.a. de interesse**

X = Número de defeitos occoridos

- **Probabilidade pedida**

$$\begin{aligned} P(X=3|X\geq 3) &= \frac{P(X=3, X\geq 3)}{P(X\geq 3)} \\ &= \frac{P(X=3)}{P(X\geq 3)} \\ &= \frac{1/10}{1-3/10} \\ &= \frac{1}{7}. \end{aligned}$$

(b) Indique a mediana (ou classe mediana) do número de defeitos findo o processo.

Classe mediana $[4,5]$

Pergunta 3 2 valores

Um armazém recebe baterias de três fábricas distintas (A, B e C) responsáveis por 20%, 30% e 50% das baterias em *stock*, respetivamente. Admita que os tempos até falha (X, em milhares de horas) das baterias provenientes das fábricas A, B e C são variáveis aleatórias independentes com funções de densidade de probabilidade $f_A(x) = \frac{1}{4}\,e^{-x/4}$, $f_B(x) = \frac{1}{3}\,e^{-x/3}$ e $f_C(x) = \frac{1}{2}\,e^{-x/2}$, para $x > 0$. Foi retirada ao acaso uma bateria do *stock* do armazém e ensaiou-se a duração dessa bateria, verificando-se que foi inferior a 1000h. Qual a probabilidade dessa bateria ser proveniente da fábrica C?

- **V.a. de interesse**

X_i : v.a. que representa a duração da bateria proveniente da fábrica $i = A, B, C$

Sabe-se que: $X_A \sim$ exponencial($\frac{1}{4}$), $X_B \sim$ exponencial($\frac{1}{3}$) e $X_C \sim$ exponencial($\frac{1}{2}$)

Sejam os acontecimentos:

  - A: “bateria fabricada na fábrica A”, com $P(A) = 0.2$
  - B: “bateria fabricada na fábrica”, com $P(B) = 0.3$
  - C: “bateria fabricada na fábrica”, com $P(C) = 0.5$

- **Probabilidade pedida**

$$\begin{aligned} P\,(\mathrm{C}|X<1) &= \frac{P\,(C \wedge X<1)}{P\,(X<1)} = \frac{P(X<1|C)\;P(C)}{P(X<1|A)\;P(A)+P(X<1|B)\;P(B)+P(X<1|C)\;P(C)} \\ &= \frac{F_{X_C}(1)\;0.5}{F_{X_A}(1)\;0.2+F_{X_B}(1)\;0.3+F_{X_C}(1)\;0.5} \\ &= \frac{(1-e^{-1/2})\;0.5}{(1-e^{-1/4})\;0.2+(1-e^{-1/3})\;0.3+(1-e^{-1/2})\;0.5} \\ &\approxeq \frac{0.196735}{0.0442398+0.0850406+0.196735} \\ &\approxeq 0.326015. \end{aligned}$$

Pergunta 4	2 valores

Seja X a variável aleatória que representa a quantidade semanal de alumínio que uma fábrica de produção de cadeiras de rodas recebe e, seja Y, a quantidade desse alumínio consumido semanalmente na produção das mesmas. Sabe-se que

$$f_{Y|X=x}(y) = \begin{cases} 3y^2x^{-3}, & y \in (0,x) \\ 0, & \text{outros casos} \end{cases}$$

com $x \in (0,1)$, e

$$f_X(x) = \begin{cases} 5x^4, & x \in (0,1) \\ 0, & \text{outros casos} \end{cases}$$

Calcular a variância da quantidade semanal de alumínio que fica por consumir.

Seja $Z = X - Y$ a variável aleatória de representa a quantidade semanal de alumínio que fica por consumir. Então,

$$V(X-Y) = V(X) + V(Y) - 2Cov(X,Y).$$

- **Cálculo da distribuição conjunta do par aleatório** (X,Y)

$$\begin{aligned} f_{X,Y}(x,y) &= f_{Y|X=x}(y) f_X(x) \\ &= 3y^2x^{-3}5x^4, \end{aligned}$$

pelo que

$$f_{X,Y}(x,y) = \begin{cases} 15y^2x & 0<y<x<1 \\ 0 & \text{outros casos} \end{cases}$$

- **Cálculo da distribuição marginal de** Y

$$\begin{aligned} f_Y(y) &= \int_y^1 15y^2x\,dx \\ &= \frac{15}{2}y^2(1-y^2), \end{aligned}$$

pelo que

$$f_Y(y) = \begin{cases} \frac{15}{2}y^2(1-y^2) & y \in (0,1) \\ 0 & \text{outros casos} \end{cases}$$

- **Cálculo da** $V(X)$ **e** $V(Y)$

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 \\ &= \int_0^1 5x^6\,dx - \left(\int_0^1 5x^5\,dx\right)^2 \\ &= \frac{5}{7} - \frac{25}{36} \\ &= 0.0198. \end{aligned}$$

$$\begin{aligned} V(Y) &= E(Y^2) - E(Y)^2 \\ &= \int_0^1 \frac{15}{2}y^4(1-y^2)\,dy - \left(\int_0^1 \frac{15}{2}y^3(1-y^2)\,dy\right)^2 \\ &= 0.4285 - 0.3906 \\ &= 0.0379. \end{aligned}$$

- **Cálculo da** $Cov(X,Y)$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 0.5357 - 0.8333 * 0.6250 = 0.0148875$$

$$E(X) = \int_0^1 5x^5 dx = \frac{5}{6}, \quad E(Y) = \int_0^1 \frac{15}{2} y^3(1-y^2) dy = \frac{15}{2}\left(\frac{1}{4} - \frac{1}{6}\right)$$

$$\begin{aligned} E(XY) &= \int_0^1 \int_0^x xy15y^2 x dy dx \\ &= \int_0^1 \int_0^x 15y^3 x^2 dy dx \\ &= \int_0^1 \frac{15}{4} x^6 dx \\ &= \frac{15}{4} \times \frac{1}{7} \\ &= \frac{15}{28} \\ &= 0.5357. \end{aligned}$$

$$\begin{aligned} V(X-Y) &= 0.0198 + 0.0379 - 2 * 0.0148875 \\ &= 0.027925. \end{aligned}$$

Pergunta 5	2 valores

Considere que o número diário de veículos detetados em excesso de velocidade por um determinado radar tem distribuição de Poisson de parâmetro 0.9. Considere que se escolhem ao acaso e de forma independente 49 dias onde são feitos os registos de excesso de velocidade por esse radar. Calcule a probabilidade aproximada do número médio de veículos detetados nessa amostra aleatória ser inferior ou igual a 1.

- **V.a. de interesse** X_i

  Seja X_i número de veículos detetados em excesso de velocidade no i-éssimo dia, $i = 1, \ldots, n$, $n = 49$.

- **Distribuição** $X_i \sim \text{poisson}(\lambda)$, $i = 1, \ldots, n$, são variáveis aleatórias iid.

- **Valor esperado de** X_i
  $\text{E}(X_i) = \lambda = 0.9$.

- **Variância de** X_i
  $\text{Var}(X_i) = \lambda = 0.9$.

- **V.a. de interesse**
  $\bar{X}_n = \sum_{i=1}^n \frac{X_i}{n}$.

- **Valor esperado de** $\bar{X}_n$
  $\text{E}(\bar{X}_n) = \text{E}(\sum_{i=1}^n X_i)/n = \sum_{i=1}^n \text{E}(X_i)/n = \lambda = 0.9$.

- **Variância de** $\bar{X}_n$
  $\text{Var}(\bar{X}_n) = \text{Var}(\sum_{i=1}^n X_i/n) = \sum_{i=1}^n \text{Var}(X_i)/n^2 = n\lambda/n^2 = \lambda/n = 0.9/49 = 0.01836735$.

- **Distribuição aproximada de $\bar{X}_n$**
  De acordo com o teorema do limite central, temos

$$\frac{\bar{X}_n - \mathrm{E}(\bar{X}_n)}{\sqrt{\mathrm{Var}(\bar{X}_n)}} = \frac{\bar{X}_n - \lambda}{\sqrt{\lambda/n}} \overset{a}{\sim} \text{normal}(0,1)$$

- **Probabilidade pedida (valor aproximado)**

$$\begin{aligned} P\left(\bar{X}_n \leq 1\right) &= P\left(\frac{\bar{X}_n - \lambda}{\sqrt{\lambda/n}} \leq \frac{1-\lambda}{\sqrt{\lambda/n}}\right) \\ &= P\left(\frac{\bar{X}_n - 0.9}{\sqrt{0.9/49}} \leq \frac{1-0.9}{\sqrt{0.9/49}}\right) \\ &\simeq \Phi(0.74) \\ &= 0.7704. \end{aligned}$$

Pergunta 6	2 valores

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

$$f_X(x;\theta) = \begin{cases} \frac{\theta^5}{24} x^4 e^{-\theta x}, & x > 0 \\ 0, & \text{outros casos} \end{cases}$$

com $\theta > 0$ desconhecido. A concretização de uma amostra aleatória de dimensão 3 proveniente de X conduziu a $x_1 = 2.5$, $x_2 = 3.5$ e $x_3 = 5.6$. Obtenha a estimativa de máxima verosimilhança de θ.

- 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 θ**
  Passo 1 - Função de verosimilhança

$$\begin{aligned} L(\theta|\underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\overset{X_i\ indep}{=} \prod_{i=1}^{3} f_{X_i}(x_i) \\ &\overset{X_i \sim X}{=} \prod_{i=1}^{3} f_{X_i}(x_i) \\ &= \prod_{i=1}^{3} \frac{\theta^5}{24} x_i^4 e^{-\theta x_i} \\ &= \frac{\theta^{15}}{24^3} \left(\prod_{i=1}^{3} x_i^4\right) e^{-\theta \sum_{i=1}^{3} x_i}. \end{aligned}$$

  Passo 2 - Função de log-verosimilhança

$$\ln L(\theta|\underline{x}) = 15\ln(\theta) - 3\ln(24) + \ln\left(\prod_{i=1}^{3} x_i^4\right) - \theta \sum_{i=1}^{3} x_i.$$

Passo 3 - Maximização

A estimativa de MV de θ é doravante representada por $\hat{\theta}$ e

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

$$\left.\frac{d \ln L(\theta|\underline{x})}{d\theta}\right|_{\theta=\hat{\theta}} = 0 \Leftrightarrow \frac{15}{\hat{\theta}} - \sum_{i=1}^{3} x_i = 0 \Rightarrow \hat{\theta} = \frac{15}{\sum_{i=1}^{3} x_i} = 1.2931.$$

$$\left.\frac{d^2 \ln L(\theta|\underline{x})}{d\theta^2}\right|_{\theta=\hat{\theta}} = -\frac{15}{1.2931^2} < 0, \text{ (proposição verdadeira)}$$

Pergunta 7 2 valores

Numa investigação sobre a incidência de uma doença que afeta animais em explorações pecuárias pretende-se estimar a proporção de animais infetados na população. Num estudo-piloto foi obtido o intervalo de confiança $[0.0336, 0.2064]$ para essa proporção.

Determine o tamanho mínimo de uma nova amostra que garanta que a amplitude do intervalo de confiança para a proporção de animais infetados não excede 0.1 ao nível de confiança aproximado de 92%.

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

$$\frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} \text{normal}(0, 1).$$

- **Obtenção dos quantis de probabilidade**

$$\Phi^{-1}(0.96) = 1.7507.$$

- **Inversão da desigualdade**

$$P\left(-1.7507 \leq \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \leq 1.7507\right) \approx 0.92 \iff P\left(\bar{X} - 1.7507\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} + 1.7507\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) \approx 0.92.$$

- **Intervalo aleatório de confiança**

$$\text{IAC}_{\approx 0.92\%}(p) = \left[\bar{X} - 1.7507\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}, \bar{X} + 1.7507\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right]$$

- **Tamanho da amostra**

$$\text{Amplitude} = 2 \times 1.7507\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq 0.1 \iff n \geq \left(\frac{2 \times 1.7507}{0.1}\right)^2 \bar{X}(1-\bar{X}).$$

Assim, por exemplo, $n_{min} = 130$ se $\bar{x} = 0.12$.

Pergunta 8	2 valores

Numa grande cidade foram inquiridos 150 adultos (escolhidos ao acaso), tendo 54 afirmado que viam o telejornal todos os dias. Teste a hipótese de que a proporção de adultos (dessa cidade) que vêem o telejornal todos os dias seja igual a 40%, ou, se pelo contrário, a propoção é inferior a este valor, decidindo com base no valor-p.

- **V.a. de interesse**

  X= indicador relativo à visualização do telejornal=$\begin{cases} 1, & \text{vê telejornal} \\ 0, & \text{outros casos} \end{cases}$

- **Distribuição**

  X v.a. com distribuição Bernoulli.

- **Hipóteses**

$$H_0 : p = p_0 = 0.40$$
$$H_1 : p = p_0 < 0.40$$

- **Estatística de teste**

$$T = \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{a}{\sim}_{H_0} \text{normal}(0,1).$$

- **Decisão**

  Atendendo a que o valor observado da estatística de teste:

$$t = \frac{54/150 - 0.4}{\sqrt{0.4 \times (1-0.4)/150}} = -1,$$
$$\text{valor-p} \approx \Phi(-1) = 1 - 0.8413 = 0.1587,$$

  pelo que H_0 não é rejeitada para os níveis de significância usuais.

Pergunta 9	2 valores

Uma empresa comercializa embalagens contendo 4 azulejos, e pretende averiguar a qualidade da sua produção. Decide então inspecionar 213 embalagens, contando em cada um das embalagens o número de azulejos que não passam no controlo de qualidade (classificados como impróprios), devido a defeitos existentes. Foram obtidos os seguintes valores:

No. azulejos impróprios	0	1	≥ 2
No. embalagens	141	48	24

Será que os dados corroboram a hipótese de que o número de azulejos impróprios por embalagem segue uma distribuição binomial de parâmetros 4 e 0.15, ao nível de significância de 1%?

- **Variável aleatória de interesse**
  X= número de azulejos impróprios por embalagem.

- **Distribuição**
  Seja $p_i = P(X = i)$ e $p_i^0 = \binom{4}{i} 0.15^i (1-0.15)^{4-i}$, para $i \in \{0,1,2,3,4\}$.

- **Hipóteses**

$$H_0 : p_i = p_i^0, i \in \{0,1,2,3,4\}$$
$$H_1 : p_i \neq p_i^0, \text{ para algum } i$$

- **Nível de significância**
  $\alpha_0 = 1\%$

- **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 é o número de classes, e O_i (E_i) a frequência observada (esperada) da classe i.

- **Cálculo das freq. esperadas**

$$p_0^0 = 0.5220,\ p_1^0 = 0.3684,\ p_2^0 = 1-(0.520+0.3684) = 0.1096$$

  pelo que

$$e_0^0 = 111.186,\ e_1^0 = 78.4692,\ e_2^0 = 23.3448.$$

- **Valor observado da estatística de teste:**

$$\begin{aligned} t &= \frac{(141-111.186)^2}{111.186} + \frac{(48-78.4692)^2}{78.4692} + \frac{(24-23.3448)^2}{23.3448} \\ &= 7.9944 + 11.83104 + 0.0183 \\ &= 19.84374. \end{aligned}$$

- **Região de rejeição de** H_0: (c,∞) com $c = F^{(-1)}_{\chi^2_{3-1}}(1-0.01) = 9.21$.

- **Decisão** Como o valor observado da estatística de teste pertence à região de rejeição, rejeita-se H_0.

Pergunta 10	2 valores

Uma nutricionista está a investigar a relação entre o índice de colesterol total (x, em miligramas por decilitro) e a tensão arterial diastólica (Y, em mmHg). Numa amostra de 10 utentes de um centro de saúde, obtiveram-se os seguintes resultados:

$$\sum_{i=1}^{10} x_i = 2459,\ \sum_{i=1}^{10} x_i^2 = 620155,\ \sum_{i=1}^{10} y_i = 820,\ \sum_{i=1}^{10} y_i^2 = 69204,\ \sum_{i=1}^{10} x_i y_i = 206055,$$

onde $[\min_{i=1,\ldots,10}(x_i), \max_{i=1,\ldots,10}(x_i)] = [200, 310]$. 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

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{10} x_i y_i - 10 \times \bar{x}\bar{y}}{\sum_{i=1}^{10} x_i^2 - 10 \times \bar{x}^2} \\
&= \frac{206055 - 10 \times 245.9 \times 82}{620155 - 10 \times 245.9^2} \\
&= 0.2852.
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&= 82 - 0.2852 \times 245.9 \\
&= 11.8671.
\end{aligned}$$

Coeficiente de determinação

$$\begin{aligned}
r^2 &= \frac{(\sum_{i=1}^{10} x_i y_i - 10 \times \bar{x}\bar{y})^2}{(\sum_{i=1}^{10} x_i^2 - 10 \times \bar{x}^2)(\sum_{i=1}^{10} y_i^2 - 10 \times \bar{y}^2)} \\
&= \frac{19509889}{15486.9 \times 1964} \\
&= 0.64142.
\end{aligned}$$

Interpretação do coeficiente de determinação

Cerca de 64% da variação total da variável resposta Y é explicada pela variável x, através do modelo de regressão linear simples ajustado.