DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA	**Probabilidades e Estatística** **LEC, LEFT, LEGI, LMAC**	1º Semestre – 2021/2022 24/02/2022 – **10:30**

Duração: **120** minutos **Exame 2**

Justifique convenientemente todas as respostas

Pergunta 1 2 valores

Uma empresa efetuou um estudo sobre a implantação de ciclovias numa cidade. Metade dos residentes declarou estar satisfeita, um sexto está parcialmente satisfeito e os restantes residentes declararam estar insatisfeitos com as ciclovias. Entre os residentes que declararam estar satisfeitos, 70% são ciclistas; entre os parcialmente satisfeitos, 20% são ciclistas; e entre os insatisfeitos, 10% são ciclistas.

Calcule a probabilidade de um residente, selecionado ao acaso e que tenha declarado não ser ciclista, estar insatisfeito com a implantação de ciclovias.

- **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
S = satisfeito com implantação de ciclovias	$P(S) = \frac{1}{2}$
PS = parcialmente satisfeito com implantação de ciclovias	$P(PS) = \frac{1}{6}$
I = insatisfeito com implantação de ciclovias	$P(I) = 1 - P(S) - P(PS) = \frac{1}{3}$
C = ciclista	$P(C) = ?$
	$P(C \mid S) = 0.7$
	$P(C \mid PS) = 0.2$
	$P(C \mid I) = 0.1$

- **Prob. pedida**

  Ao aplicar o teorema de Bayes, obtemos

$$\begin{aligned}
P(I \mid \bar{C}) &= \frac{P(\bar{C} \mid I) \times P(I)}{P(\bar{C})} \\
&= \frac{[1 - P(C \mid I)] \times P(I)}{1 - [P(C \mid S) \times P(S) + P(C \mid PS) \times P(PS) + P(C \mid I) \times P(I)]} \\
&= \frac{(1-0.1) \times \frac{1}{3}}{1 - \left(0.7 \times \frac{1}{2} + 0.2 \times \frac{1}{6} + 0.1 \times \frac{1}{3}\right)} \qquad \left[= \frac{0.9 \times \frac{1}{3}}{1 - \frac{5}{12}} = \frac{18}{35}\right] \\
&\simeq 0.514286.
\end{aligned}$$

Pergunta 2 2 valores

Uma fábrica de componentes elétricas adquire a um fornecedor lotes de fusíveis dos quais 20% são defeituosos. Admita que, para verificar a qualidade de cada lote, a responsável pelo controlo de qualidade da fábrica recolhe uma amostra, ao acaso e com reposição, de 14 fusíveis. Seja X a variável aleatória que representa o número de fusíveis defeituosos nesta amostra. Tal responsável decide não aceitar o lote se o número de fusíveis defeituosos na amostra for superior à mediana de X.

Obtenha a mediana de X; além disso, calcule a probabilidade de um lote ser aceite, sabendo que a amostra respetiva contém pelo menos um fusível defeituoso.

- **V.a.**

  X = número de fusíveis defeituosos, em 14 selecionados ao acaso e com reposição de um lote

- **Distribuição**

  $X \sim \text{binomial}(n,p)$

  $n = 14$

  $p = P(\text{selecionar um fusível defeituoso do lote}) = 0.2$

- **F.p. de X**

  $P(X=x) = \binom{14}{x}\, 0.2^x\, (1-0.2)^{14-x}, \quad x = 0,1,\ldots,14$

- **Mediana de X**

$$
\begin{aligned}
me = me(X) \in \{0,1,\ldots,14\} \quad : \quad & \frac{1}{2} \leq F_X(me) \leq \frac{1}{2} + P(X = me) \\
& F_X(me^-) \leq \frac{1}{2} \leq F_X(me)
\end{aligned}
$$

  A consulta das tabelas da f.d. da binomial leva-nos a concluir que

$$
0.4481 = F_X(2) = F_X(3^-) \leq \frac{1}{2} \leq F_X(3) = 0.6982,
$$

  pelo que $me = me(X) = 3$.

- **Probabilidade pedida**

$$
\begin{aligned}
P(\text{aceitar lote} \mid X \geq 1) &= P(X \leq me \mid X \geq 1) \\
&= \frac{P(X \leq me, X \geq 1)}{P(X \geq 1)} \\
&= \frac{P(1 \leq X \leq 3)}{1 - P(X < 1)} \\
&= \frac{P(0 < X \leq 3)}{1 - P(X = 0)} \\
&= \frac{F_X(3) - F_X(0)}{1 - F_X(0)} \\
&\overset{tabelas/calc.}{=} \frac{0.6982 - 0.0440}{1 - 0.0440} \\
&= 0.684305.
\end{aligned}
$$

Pergunta 3	2 valores

A vida útil (X, em anos), de determinada marca de televisores é normalmente distribuída com desvio padrão igual a 3 anos. Sabe-se ainda que 2.5% desses televisores duram no máximo 5 anos.

Após ter determinado o valor esperado de X, obtenha $E(0.5\,X^2 + 1.5\,X)$.

- **V.a.**

  $X =$ vida útil de televisor de certa marca

  $X \sim \text{normal}(\mu, \sigma^2 = 3^2)$

- **Cálculo de $\mu = E(X)$**

$$
\begin{aligned}
\mu \quad : \quad & F_X^{-1}(0.025) = 5 \\
& P(X \leq 5) = 0.025
\end{aligned}
$$

$$
\begin{aligned}
\mu \quad : \quad & \Phi\left(\frac{5-\mu}{\sigma}\right) = 0.025 \\
& \frac{5-\mu}{3} = \Phi^{-1}(0.025) \\
& \frac{5-\mu}{3} = -1.9600 \qquad \text{(quantil tabelado)} \\
& \mu = 5-(-1.9600)\times 3 \\
& \mu = 10.88
\end{aligned}
$$

- **Valor esperado pedido**

$$
\begin{aligned}
E(0.5X^2+1.5X) &= 0.5\times E(X^2)+1.5\times E(X) \\
&= 0.5\times[V(X)+E^2(X)]+1.5\times E(X) \\
&= 0.5\times(\sigma^2+\mu^2)+1.5\times\mu \\
&= 0.5\times(9+10.88^2)+1.5\times 10.88 \\
&= 80.0072
\end{aligned}
$$

Pergunta 4	2 valores

Num *stand*, sempre que é vendido um automóvel, o comprador é aliciado a adquirir um seguro de extensão da garantia. A função de probabilidade conjunta do número mensal de automóveis vendidos (X) e do número mensal de seguros de extensão da garantia adquiridos (Y) é a seguinte:

	Y		
X	0	1	2
0	0.3	0	0
1	0.1	0.3	0
2	0	0.1	0.2

Determine o coeficiente de correlação entre X e Y e comente o valor obtido.

- **Par aleatório e f.p. marginais**

  (X,Y)

  $X =$ número mensal de automóveis vendidos

  $Y =$ número mensal de seguros de extensão da garantia adquiridos

- **Correlação pedida**

$$
corr(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)\,V(Y)}} = \frac{E(XY)-E(X)\times E(Y)}{\sqrt{V(X)\,V(Y)}}.
$$

  Logo são necessários alguns cálculos auxiliares que envolverão as f.p. conjunta de (X,Y) e marginais de X e Y dadas por $P(X=x)=\sum_y P(X=x,Y=y)$ e $P(Y=y)=\sum_x P(X=x,Y=y)$, respectivamente.

	Y			
X	0	1	2	$P(X=x)$
0	0.3	0	0	0.3
1	0.1	0.3	0	0.4
2	0	0.1	0.2	0.3
$P(Y=y)$	0.4	0.4	0.2	1

- **Valor esperado e variância de X**

$$\begin{aligned}
E(X) &= \sum_{x=0}^{2} x \times P(X=x) \\
&= 1 \times 0.4 + 2 \times 0.3 \\
&= 1 \\
V(X) &= E(X^2) - E^2(X) \\
&= \sum_{x=0}^{2} x^2\, P(X=x) - E^2(X) \\
&= \left(1^2 \times 0.4 + 2^2 \times 0.3\right) - 1^2 \\
&= 1.6 - 1 \\
&= 0.6
\end{aligned}$$

- **Valor esperado e variância de Y**

$$\begin{aligned}
E(Y) &= \sum_{y=0}^{2} y \times P(Y=y) \\
&= 1 \times 0.4 + 2 \times 0.2 \\
&= 0.8 \\
V(Y) &= E(Y^2) - E^2(Y) \\
&= \sum_{y=0}^{2} y^2\, P(Y=y) - E^2(Y) \\
&= \left(1 \times 0.4 + 2^2 \times 0.2\right) - 0.8^2 \\
&= 1.2 - 0.64 \\
&= 0.56
\end{aligned}$$

- **Valor esperado de XY**

$$\begin{aligned}
E(XY) &= \sum_{x=0}^{2}\sum_{y=0}^{2} xy \times P(X=x, Y=y) \\
&= 1 \times 1 \times 0.3 + 2 \times 1 \times 0.1 + 2 \times 2 \times 0.2 \\
&= 1.3
\end{aligned}$$

- **Covariância entre X e Y**

$$\begin{aligned}
cov(X,Y) &= E(XY) - E(X) \times E(Y) \\
&= 1.3 - 1 \times 0.8 \\
&= 0.5
\end{aligned}$$

- **Correlação pedida** (cont.)

$$\begin{aligned}
corr(X,Y) &= \frac{cov(X,Y)}{\sqrt{V(X)\,V(Y)}} \\
&= \frac{0.5}{\sqrt{0.6 \times 0.56}} \\
&\simeq 0.862582
\end{aligned}$$

- **Comentários**
  - [É sabido que: caso X e Y sejam v.a. independentes, então $corr(X,Y) = 0$.]
    Uma vez que $corr(X,Y) \neq 0$, concluímos que X e Y são v.a. dependentes.
  - Dado que $corr(X,Y) > 0$ podemos adiantar que X e Y tenderão a variar no mesmo sentido relativamente aos respectivos valores esperados.
  - Como $|corr(X,Y)| \simeq 0.86$ pouco dista de 1, pelo que as v.a. estão fortemente correlacionadas.

Pergunta 5	2 valores

Admita que as oscilações no preço (em euros) das ações de uma empresa a cada minuto são variáveis aleatórias independentes e identicamente distribuídas à variável aleatória X com função de probabilidade

$$P(X=x)=\begin{cases}0.5, & x=-0.05\\ 0.2, & x=0\\ 0.3, & x=0.05\\ 0, & x\neq -0.05,0,0.05\end{cases}$$

e variância $V(X)=0.0019$.

Calcule a probabilidade aproximada de em três horas a oscilação total no preço das ações da empresa ser superior a 15 cêntimos.

- **V.a.; valor esperado, variância e terceiro quartil comuns**

  X_i = oscilação no preço das ações da empresa no minuto i, $\quad i=1,\ldots,n$

  $n=180$ minutos ou 3 horas

  $X_i \overset{i.i.d.}{\sim} X$

  $E(X_i)=E(X)=\mu=\sum_x x\times P(X=x)=(-0.05)\times 0.5+0\times 0.2+0.05\times 0.3=-0.01$

  $V(X_i)=V(X)=\sigma^2 \overset{enunc.}{=} 0.0019$

- **V.a. de interesse**

  $S_n=\sum_{i=1}^n X_i$ = oscilação no preço das ações da empresa em n minutos

- **Valor esperado e variância de S_n**

  $E(S_n)=E\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n E(X_i) \overset{X_i\sim X}{=} n\,E(X)=n\,\mu$

  $V(S_n)=V\left(\sum_{i=1}^n X_i\right) \overset{X_i\ indep.}{=} \sum_{i=1}^n V(X_i) \overset{X_i\sim X}{=} n\,V(X)=n\,\sigma^2$

- **Distribuição aproximada de S_n**

  Segundo o teorema do limite central (TLC),

  $$\frac{S_n-E(S_n)}{\sqrt{V(S_n)}}=\frac{S_n-n\,E(X)}{\sqrt{n\,V(X)}}=\frac{S_n-n\,\mu}{\sqrt{n\,\sigma^2}} \overset{a}{\sim} \text{normal}(0,1).$$

- **Probabilidade pedida (valor aproximado)**

  $$\begin{aligned}P(S_n>0.15) &= 1-P\left(\frac{S_n-n\mu}{\sqrt{n\sigma^2}}\leq\frac{0.15-n\mu}{\sqrt{n\sigma^2}}\right)\\ &\overset{TLC}{\simeq} 1-\Phi\left(\frac{0.15-n\mu}{\sqrt{n\sigma^2}}\right)\\ &= 1-\Phi\left(\frac{0.15-180\times(-0.01)}{\sqrt{180\times 0.0019}}\right)\\ &\simeq 1-\Phi(3.33)\\ &\overset{tabelas/calc.}{=} 1-0.999566=0.000434.\end{aligned}$$

Pergunta 6	2 valores

Admita que a quantidade de água disponível num reservatório, com capacidade máxima de 1000 L, é representada pela variável aleatória X (em milhares de litros) com função de densidade de probabilidade

$$f_X(x)=\begin{cases}2\,\beta\,x\,(1-x^2)^{\beta-1}, & x\in(0,1)\\ 0, & \text{caso contrário,}\end{cases}$$

onde β é um parâmetro positivo desconhecido.

Determine a estimativa de máxima verosimilhança de $P(X > 0.5)$, baseada na amostra $(x_1, \ldots, x_6)$ proveniente da população X e tal que $\sum_{i=1}^{6} \ln\left(1 - x_i^2\right) \simeq -3.698$.

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

  X = quantidade de água disponível num reservatório (em milhares de litros)

  $f_X(x) = \begin{cases} 2\,\beta\,x\,(1-x^2)^{\beta-1}, & x \in (0,1) \\ 0, & \text{caso contrário,} \end{cases}$

- **Parâmetro desconhecido**

  $\beta \quad (\beta > 0)$

- **Amostra**

  $\underline{x} = (x_1, \ldots, x_n)$, amostra de dimensão $n = 6$, proveniente da população X e tal que $\sum_{i=1}^{6} \ln(1 - x_i^2) \simeq -3.698$.

- **Obtenção da estimativa de MV de β**

  Passo 1 — Função de verosimilhança

$$
\begin{aligned}
L(\beta \mid \underline{x}) &= f_{\underline{X}}(\underline{x}) \\
&\overset{X_i\ indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \\
&\overset{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) \\
&= \prod_{i=1}^{n} \left[2\,\beta\,x_i\,(1-x_i^2)^{\beta-1}\right] \\
&= 2^n\,\beta^n \times \left(\prod_{i=1}^{n} x_i\right) \times \left(\prod_{i=1}^{n}(1-x_i^2)\right)^{\beta-1}, \quad \beta > 0
\end{aligned}
$$

  Passo 2 — Função de log-verosimilhança

$$
\ln L(\beta|\underline{x}) = n\ln(2) + n\ln(\beta) + \sum_{i=1}^{n}\log(x_i) + (\beta-1)\sum_{i=1}^{n}\ln\left(1-x_i^2\right), \quad \beta > 0
$$

  Passo 3 — Maximização

  A estimativa de MV de β é doravante representada por $\hat{\beta}$ e

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

$$
\begin{cases} \frac{n}{\hat{\beta}} + \sum_{i=1}^{n}\ln\left(1-x_i^2\right) = 0 \\ -\frac{n}{\hat{\beta}^2} < 0 \quad \text{(prop. verdadeira)} \end{cases}
$$

  Passo 4 — Estimativa de MV de β

$$
\begin{aligned}
\hat{\beta} &= -\frac{n}{\sum_{i=1}^{n}\ln\left(1-x_i^2\right)} \\
&\simeq -\frac{6}{-3.698} \\
&\simeq 1.622499
\end{aligned}
$$

- **Outro parâmetro desconhecido**

$$\begin{aligned} h(\beta) &= P(X > 0.5) \\ &= \int_{0.5}^{+\infty} f_X(x)\,dx \\ &= \int_{0.5}^{1} 2\,\beta\,x\,\left(1-x^2\right)^{\beta-1}\,dx \\ &= -\left(1-x^2\right)^{\beta}\Big|_{0.5}^{1} \\ &= 0.75^{\beta} \end{aligned}$$

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

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

$$\begin{aligned} \widehat{h(\beta)} &= h(\hat{\beta}) \\ &= 0.75^{\hat{\beta}} \\ &\simeq 0.75^{1.622499} \\ &\simeq 0.627028. \end{aligned}$$

Pergunta 7 2 valores

A duração (em segundos) de certa experiência em laboratório pode ser considerada uma variável aleatória X com distribuição normal de valor esperado e variância desconhecidos. Com base em 40 experiências distintas e consideradas independentes, obtiveram-se a média amostral e a variância amostral corrigida seguintes: $\bar{x} = 37.5$ e $s^2 = 8.76$.

Determine um intervalo de confiança a 90% para o desvio padrão de X.

- **V.a. de interesse**

  X = duração (em segundos) da experiência

- **Situação**

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

  μ desconhecido

  σ^2 DESCONHECIDO

- **Obtenção de IC para σ**

  Passo 1 — Selecção da v.a. fulcral para σ

$$Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

  [dado que é suposto determinar um IC para o desvio padrão de uma população normal, com valor esperado desconhecido.]

  Passo 2 — Obtenção dos quantis de probabilidade

  Dado que $n = 40$ e $(1-\alpha)\times 100\% = 90\%$, usaremos os quantis

$$(a_\alpha, b_\alpha) \quad : \quad \left\{ \begin{array}{l} P(a_\alpha \leq Z \leq b_\alpha) = 1-\alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(39)}}(0.05) \stackrel{tabela/calc.}{=} 25.70 \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(39)}}(0.95) \stackrel{tabela/calc.}{=} 54.57. \end{array} \right.$$

Passo 3 — Inversão da desigualdade $a_\alpha \le Z \le b_\alpha$

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

$$P\left[a_\alpha \le \frac{(n-1)S^2}{\sigma^2} \le b_\alpha\right] = 1-\alpha$$

$$P\left[\frac{1}{b_\alpha} \le \frac{\sigma^2}{(n-1)S^2} \le \frac{1}{a_\alpha}\right] = 1-\alpha$$

$$P\left[\sqrt{\frac{(n-1)S^2}{b_\alpha}} \le \sigma \le \sqrt{\frac{(n-1)S^2}{a_\alpha}}\right] = 1-\alpha$$

Passo 4 — Concretização

Atendendo à expressão geral do IC para σ,

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

ao par de quantis acima e ao facto de $s^2 = 8.76$, temos:

$$IC_{99\%}(\sigma) \quad = \quad \left[\sqrt{\frac{(40-1)\times 8.76}{54.57}}, \sqrt{\frac{(40-1)\times 8.76}{25.70}}\right] \simeq [2.502116, 3.646009].$$

Pergunta 8	2 valores

Determinado partido político afirma que 80% da população é favorável a determinada revisão legislativa.

Numa amostra aleatória de 100 pessoas que responderam a um questionário, 69 destas mostraram-se favoráveis a tal revisão.

Efetue um teste de hipóteses adequado e obtenha o correspondente valor-p aproximado. Que decisão deve tomar ao nível de significância de 5% ?

- **V.a. de interesse**

  X = indicador de resposta favorável à revisão legislativa = $\begin{cases} 1, & \text{se a resposta é favorável} \\ 0, & c.c. \end{cases}$

- **Situação**

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

  $p = P(\text{resposta favorável à revisão legislativa})$ DESCONHECIDA

  $n = 100 > 30$ (suficientemente grande)

- **Hipóteses**

  $H_0 : p = p_0 = 0.8$

  $H_1 : p \neq p_0$

- **Estatística de teste**

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

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

  O teste é bilateral ($H_1 : p \neq p_0$), logo a região de rejeição de H_0 é do tipo $W = (-\infty, -c) \cup (c, +\infty)$.

- **Decisão (com base no valor-p)**

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

$$
\begin{aligned}
t &= \frac{\bar{x}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\
&= \frac{\frac{69}{100}-0.8}{\sqrt{\frac{0.8\,(1-0.8)}{100}}} \\
&\simeq -2.75 \\
valor-p &= 2\times P(T>|t|\mid H_0) \\
&\simeq 2\times[1-\Phi(|t|)] \\
&\simeq 2\times[1-\Phi(2.75)] \\
&\overset{tabelas/calc.}{=} 2\times(1-0.9970) \\
&= 0.006.
\end{aligned}
$$

  Consequentemente, é suposto:

  - [não rejeitar H_0 a qualquer n.s. $\alpha_0 \le valor-p = 0.6\%$;]
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > valor-p = 0.6\%$, designadamente ao n.u.s. de 5%.

Pergunta 9 — 2 valores

Seja X a variável aleatória que representa o estado de um sistema (num total de quatro estados possíveis).

Uma engenheira conjeturou a hipótese H_0 de que a função de probabilidade de X é dada por: $P(X=i) = p_i = p_i^0$, para $i=1,2,3,4$, onde $p_1^0 = 0.084$, $p_2^0 = 0.250$, $p_3^0 = 0.166$ e $p_4^0 = 0.500$.

Numa amostra casual de 60 experiências aleatórias, foram registadas as seguintes frequências para cada estado:

Estado (i)	1	2	3	4
Frequência absoluta observada (o_i)	3	27	7	23

Após ter calculado as frequências absolutas esperadas sob H_0, averigue se os dados são consistentes com H_0, ao nível de significância de 5%.

- **V.a. de interesse**

  $X =$ estado do sistema

  $$P(X=x_i) = \begin{cases} p_i, & i=1,2,3,4 \\ 0, & \text{caso contrário} \end{cases}$$

- **Hipóteses**

  $H_0: p_i = p_i^0, \quad i=1,\ldots,4 \quad$ onde $p_1^0 = 0.084$, $p_2^0 = 0.250$, $p_3^0 = 0.166$ e $p_4^0 = 0.500$

  $H_1: p_i \neq p_i^0, \quad$ para algum i

- **N.s.**

  $\alpha_0 = 5\%$

- **Estatística de teste**

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

onde:

k = número de classes = 4 (estados);

O_i = frequência absoluta observável da classe i;

E_i = frequência absoluta esperada, sob H_0, da classe i;

β = número de parâmetros a estimar = 0

[dado que a distribuição conjecturada em H_0 está completamente especificada].

- **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

$$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(4-0-1)}}(1-0.05) \stackrel{tabela/calc.}{=} 7.815.$$

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

  As frequências absolutas esperadas sob H_0 são dadas por $E_i = n \times p_i^0$, $i = 1,2,3,4$ e iguais a: $E_1 = 60 \times 0.084 = 5.04$; $E_2 = 60 \times 0.250 = 15.00$; $E_3 = 60 \times 0.166 = 9.96$; $E_4 = 60 \times 0.50 = 30.00$.

- **Decisão (com base no valor-p)**

  No cálculo do valor observado da estatística de teste convém recorrer à seguinte tabela auxiliar:

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	3	5.04	$\frac{(3-5.04)^2}{5.04} \simeq 0.826$
2	27	15.00	$\frac{(27-15.00)^2}{15.00} = 9.6$
3	7	9.96	$\frac{(7-9.96)^2}{9.96} \simeq 0.880$
4	23	30.00	$\frac{(23-30.00)^2}{30.00} \simeq 1.633$
	$\sum_{i=1}^4 o_i = n = 60$	$\sum_{i=1}^4 E_i = n = 60$	$t = \sum_{i=1}^4 \frac{(o_i-E_i)^2}{E_i} = 12.939$

  Como $t = 12.939 \in W = (7.815, +\infty)$ devemos rejeitar H_0 ao n.s. $\alpha_0 = 5\%$ (ou a qualquer n.s. superior a $\alpha_0 = 5\%$).

Pergunta 10 2 valores

O modelo de regressão linear simples foi usado para averiguar da existência de uma eventual relação entre os tempos de vida (em anos) de duas componentes A e B de um sistema. Para tal, foi medido o tempo de vida de A (x) e de B (Y) em $n = 10$ instâncias, tendo-se obtido as somas seguintes:

$$\sum_{i=1}^n x_i = 7.5, \quad \sum_{i=1}^n x_i^2 = 8.81, \quad \sum_{i=1}^n y_i = 98.4, \quad \sum_{i=1}^n y_i^2 = 1110.48, \quad \sum_{i=1}^n x_i y_i = 66.00,$$

onde $[\min_{i=1,\dots,n} x_i, \max_{i=1,\dots,n} x_i] = [0.1, 1.9]$.

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

Obtenha a reta de mínimos quadrados com base nos dados fornecidos; além disso, calcule o coeficiente de determinação e interprete o seu valor.

- **Modelo de RLS**

  Y = tempo de vida (em anos) da componente B (v.a. resposta)

$x =$ tempo de vida (em anos) da componente A (variável explicativa)

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

$E(\epsilon_i) = 0,\ V(\epsilon_i) = \sigma^2, \quad i = 1, \ldots, n$

$cov(\epsilon_i, \epsilon_j) = 0,\ i \neq j,\ i, j = 1, \ldots, n]$

- **Recta de regressão de mínimos quadrados**

  Temos

  - $n = 15$
  - $\sum_{i=1}^n x_i = 7.5$

    $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{7.5}{10} = 0.75$

    $\sum_{i=1}^n x_i^2 = 8.81$

    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 \simeq 8.81 - 10 \times 0.75^2 = 3.185$
  - $\sum_{i=1}^n y_i = 98.4$

    $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{98.4}{10} = 9.84$

    $\sum_{i=1}^n y_i^2 = 1\,110.48$

    $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 1\,110.48 - 10 \times 9.84^2 = 142.2240$
  - $\sum_{i=1}^n x_i y_i = 66.00$

    $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 66.00 - 10 \times 0.75 \times 9.84 = -7.8.$

  Logo, as estimativas de MQ dos parâmetros desconhecidos β_1 e β_0, bem como a recta de regressão de MQ, são dadas por:

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} \\
&= \frac{-7.8}{3.185} \\
&\simeq -2.448980; \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&\simeq 9.84 - (-2.448980) \times 0.75 \\
&\simeq 11.676735; \\
\hat{E}(Y \mid x) &= \hat{\beta}_0 + \hat{\beta}_1 \times x = 11.676735 + (-2.448980) \times x, \quad x \in \left[\min_{i=1,\ldots,10} x_i,\ \max_{i=1,\ldots,10} x_i\right] = [0.1\,,\,1.9]
\end{aligned}
$$

- **Cálculo do coeficiente de determinação**

$$
\begin{aligned}
r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\
&= \frac{(-7.8)^2}{3.185 \times 142.2240} \\
&\simeq 0.134310.
\end{aligned}
$$

- **Interpretação coeficiente de determinação**

  Cerca de 13% da variação total da variável resposta Y é explicada pela variável x, através do modelo de regressão linear simples ajustado, donde possamos afirmar que a recta estimada parece ajustar-se mal ao conjunto de dados.