Probabilidade e Estatística

LEC, LEFT, LEGI, LMAC

1º Semestre – 2023/2024
15/01/2024
15:30–17:30

Duração: **120** minutos **Exame de época normal**

- As questões de escolha múltipla têm uma e apenas uma resposta correta. Caso assinale uma resposta errada ou mais de uma resposta, serão descontados 0.5 valores.
- As respostas às questões de escolha múltiplas têm de ser assinaladas na folha de instruções previamente distribuída.
- Para as restantes questões, terá de justificar convenientemente as suas respostas.

Pergunta 1 2 valores

Num centro de recolha, verificou-se que 70% das peças recebidas correspondem a roupa, 20% a calçado e 10% a acessórios. Verificou-se ainda que 70% da roupa e 95% dos acessórios estavam em condições de ser revendidos, contrariamente a 75% do calçado que não estava em condições de ser revendido.

Selecionada uma peça ao acaso, qual é a probabilidade de que seja uma peça de roupa, sabendo que está em condições de ser revendida?

- **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
A = peça de roupa	$P(A) = 0.7$
B = peça de calçado	$P(B) = 0.2$
C = peça de acessório	$P(C) = 0.1$
R = peça estar em condições para revenda	
	$P(R \mid A) = 0.7$
	$P(\bar{R} \mid B) = 0.75$
	$P(R \mid C) = 0.95$

- **Prob. pedida**

  Ao invocarmos o teorema de Bayes, temos

$$
\begin{aligned}
P(A \mid R) &= \frac{P(R \mid A) \times P(A)}{P(R)} \\
&= \frac{P(R \mid A) \times P(A)}{P(R \mid A) \times P(A) + P(R \mid B) \times P(B) + P(R \mid C) \times P(C)} \\
&= \frac{0.7 \times 0.7}{0.7 \times 0.7 + (1 - 0.75) \times 0.2 + 0.95 \times 0.1} \\
&= \frac{0.49}{0.635} \\
&\simeq 0.7717.
\end{aligned}
$$

Pergunta 2 2 valores

Em certa experiência, a probabilidade de uma partícula escapar à trajetória prevista é igual a 0.03.

Calcule o quantil de probabilidade 0.95 ($q_{0.95}$) do número de vezes (X) que a partícula escapa à trajetória, em 18 realizações independentes de tal experiência. Determine $P(X \geq q_{0.95})$.

- **V.a. e f.p.**

  X = número de vezes em que a partícula escapa à trajetória, em 18 realizações da experiência

  $X \sim \text{binomial}(n = 18, p = 0.03)$

  $P(X = x) = \binom{18}{x} 0.03^x (1-0.03)^{18-x}, \quad x = 0, 1, \ldots, 18$

- **Quantil pedido**

  O quantil de probabilidade p de X, q_p $(p = 0.95)$, define-se do modo seguinte:

  $$\begin{aligned} q_p \in \{0,1,\ldots,18\} \quad : \quad & P(X \le q_p) \ge p \quad \text{e} \quad P(X \ge q_p) \ge 1-p \\ & p \le F_X(q_p) \le p + P(X = q_p) \quad \Leftrightarrow \quad F_X[(q_p)^-] \le p \le F_X(q_p). \end{aligned}$$

  De acordo com a tabela/calc., $F_X(1) = P(X \le 1) = F_X(2^-) = 0.8997$ e $F_X(2) = 0.9843$. Logo,

  $$0.8997 = F_X(2^-) \le 0.95 \le F_X(2) = 0.9843,$$

  pelo que o quantil de probabilidade 0.95 da v.a. X é igual a $q_{0.95} = 2$.

- **Probabilidade pedida**

  $$P(X \ge q_{0.95}) = P(X \ge 2) = 1 - P(X < 2) = 1 - P(X \le 1) = 1 - F_X(1) \stackrel{tabelas, calc.}{=} 1 - 0.8997 = 0.1003.$$

Pergunta 3	2 valores

Admita que o tempo (em anos) até à ocorrência do próximo terramoto em determinada região é uma variável aleatória X com distribuição exponencial com parâmetro λ $(\lambda > 0)$. Admita também que, sabendo que o último terramoto ocorreu há mais de dois anos nessa região, a probabilidade de ter de se esperar adicionalmente mais de quatro anos pela ocorrência do próximo terramoto é igual a 0.818731.

Qual é o valor de λ arredondado às centésimas? **A**: 0.03 **B**: 0.43 **C**: 0.28 **D**: 0.05

- **V.a. de interesse, f.d.p. e f.d.**

  X = tempo (em anos) até à ocorrência do próximo terramoto...

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

  $$f_X(x) \stackrel{form.}{=} \begin{cases} \lambda\, e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

  $$F_X(x) = P(X \le x) = \int_{-\infty}^x f_X(t)\,dt = \begin{cases} 0, & x < 0 \\ \int_0^x \lambda\, e^{-\lambda t}\,dt = 1 - e^{-\lambda x}, & x \ge 0 \end{cases}$$

- **Obtenção de λ**

  Ao invocarmos a propriedade da falta de memória da distribuição exponencial, obtemos

  $$\begin{aligned} \lambda > 0 \quad : \quad & P(X > 2+4 \mid X > 2) = 0.818731 \\ & P(X > 4) = 0.818731 \quad \Leftrightarrow \quad 1 - F_X(4) = 0.818731 \\ & e^{-\lambda \times 4} = 0.818731 \quad \Leftrightarrow \quad \lambda = -1/4 \times \ln(0.818731) \\ & \lambda \simeq 0.05. \end{aligned}$$

  [Em alternativa,

  $$\begin{aligned} \lambda > 0 \quad : \quad & P(X > 2+4 \mid X > 2) = 0.818731 \quad \Leftrightarrow \quad \frac{P(X > 2+4, X > 4)}{P(X > 2)} = 0.818731 \\ & \frac{P(X > 2+4)}{P(X > 2)} = 0.818731 \quad \Leftrightarrow \quad \frac{e^{-\lambda \times (2+4)}}{e^{-\lambda \times 2}} = 0.818731 \\ & e^{-\lambda \times 4} = 0.818731 \quad \Leftrightarrow \quad \lambda = -\frac{1}{4} \times \ln(0.818731) \quad \Leftrightarrow \quad \lambda \simeq 0.05.] \end{aligned}$$

Pergunta 4	2 valores

O retorno do aluguer de um equipamento em dois dias consecutivos é representado pelo par aleatório (X,Y), com função de probabilidade conjunta dada por

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

Calcule o valor esperado e a variância de $X-Y$.

- **Par aleatório**

  (X,Y)

  X = retorno do aluguer do equipamento no primeiro dia

  Y = retorno do aluguer do equipamento no segundo dia

- **V.a. de interesse**

  $X-Y$ = diferença entre os retornos do aluguer do equipamento em dois dias consecutivos

- **Valor esperado e variância pedidos**

  Uma vez que $E(X-Y)=E(X)-E(Y)$ e $V(X-Y)=V(X)+V(Y)-2\,cov(X,Y)$, são necessários alguns cálculos auxiliares que envolverão a f.p. conjunta de (X,Y) e as f.p. marginais de X e Y dadas por $P(X=x)=\sum_y P(X=x,Y=y)$ e $P(Y=y)=\sum_x P(X=x,Y=y)$.

	Y			$P(X=x)$
X	0	1	2	
0	0.1	0.2	0.1	0.4
1	0.2	0.1	0	0.3
2	0	0.3	0	0.3
$P(Y=y)$	0.3	0.6	0.1	1

- **Valores esperados e variâncias de X e de Y**

$$\begin{aligned}
E(X) &= \sum_{x=0}^{2} x \times P(X=x) = 1\times 0.3 + 2\times 0.3 = 0.9 \\
E(X^2) &= \sum_{x=0}^{2} x^2\, P(X=x) = 1^2\times 0.3 + 2^2 \times 0.3 = 1.5 \\
V(X) &= E(X^2) - E^2(X) = 1.5 - 0.9^2 = 0.69 \\
E(Y) &= \sum_{y=0}^{2} y \times P(Y=y) = 1\times 0.6 + 2\times 0.1 = 0.8 \\
E(Y^2) &= \sum_{y=0}^{2} y^2\, P(Y=y) = 1^2\times 0.6 + 2^2 \times 0.1 = 1 \\
V(Y) &= E(Y^2) - E^2(Y) = 1 - 0.8^2 = 0.36
\end{aligned}$$

- **Valor esperado de XY**

$$E(XY) = \sum_{x=0}^{2}\sum_{y=0}^{2} xy \times P(X=x,Y=y) = 1\times 1\times 0.1 + 2\times 1\times 0.3 = 0.7$$

- **Covariância entre X e Y**

$$cov(X,Y) = E(XY) - E(X)\times E(Y) = 0.7 - 0.9\times 0.8 = -0.02$$

- **Valor esperado e variância pedidos** (cont.)

$$\begin{aligned} E(X-Y) &= E(X)-E(Y) \\ &= 0.9-0.8 \\ &= 0.1 \end{aligned}$$

$$\begin{aligned} V(X-Y) &= V(X)+V(Y)-2\,cov(X,Y) \\ &= 0.69+0.36-2\times(-0.02) \\ &= 1.09. \end{aligned}$$

Pergunta 5	2 valores

Numa linha de montagem, o número de unidades rejeitadas diariamente pode ser descrito por uma variável aleatória X com distribuição geométrica com parâmetro $p = 0.05$. Admita que os números de unidades rejeitadas em dias sucessivos são variáveis aleatórias independentes e identicamente distribuídas a X.

Obtenha um valor aproximado para a probabilidade de o número total de unidades rejeitadas em 40 dias ser superior a 780.

- **V.a.; valor esperado e variância comuns**

  X_i = número de unidades rejeitadas no dia i, $\quad i = 1, \ldots, n$

  $n = 40$

  $X_i \overset{indep}{\sim}$ geométrica$(p = 0.05)$

  $E(X_i) = E(X) = \mu \overset{form.}{=} 1/p = 20$

  $V(X_i) = V(X) = \sigma^2 \overset{form.}{=} (1-p)/p^2 = 380$

- **V.a. de interesse**

  $S_n = \sum_{i=1}^n X_i$ = número total de unidades rejeitadas em n dias

- **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**

  De acordo com o teorema do limite central (TLC),

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

- **Prob. pedida** (valor aproximado)

$$\begin{aligned} P\left[S_n > 780\right] &= 1 - P\left(\frac{S_n - n\,\mu}{\sqrt{n\,\sigma^2}} \le \frac{780 - n\,\mu}{\sqrt{n\,\sigma^2}}\right) \\ &\overset{TLC}{\simeq} 1 - \Phi\left(\frac{780 - n\,\mu}{\sqrt{n\,\sigma^2}}\right) \\ &= 1 - \Phi\left(\frac{780 - 40\times 20}{\sqrt{40\times 380}}\right) \\ &\simeq 1 - \Phi(-0.16) \\ &= \Phi(0.16) \\ &\overset{tabelas/calc.}{\simeq} 0.5636. \end{aligned}$$

Pergunta 6	2 valores

Admita que a fração de carga máxima suportada por uma estrutura em relação à carga máxima possível é representada pela variável aleatória X com função de densidade de probabilidade dada por

$$f_X(x) = \begin{cases} 3\beta x^2(1-x^3)^{\beta-1}, & x \in (0,1) \\ 0, & \text{outros valores de } x, \end{cases}$$

onde β é um parâmetro desconhecido positivo.

Deduza o estimador de máxima verosimilhança de β, com base numa amostra aleatória de dimensão n proveniente de X.

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

  X = fração de carga máxima suportada por uma estrutura em relação à carga máxima possível

  $$f_X(x) = \begin{cases} 3\beta x^2(1-x^3)^{\beta-1}, & x \in (0,1) \\ 0, & \text{outros valores de } x \end{cases}$$

- **Parâmetro desconhecido**

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

- **Amostra; amostra aleatória**

  $(x_1, \ldots, x_n)$, amostra de dimensão n proveniente de X.

  $(X_1, X_2, \ldots, X_n)$, amostra aleatória de dimensão n proveniente de X.

- **Dedução do estimador de MV de β**

  Passo 1 — Função de verosimilhança

$$\begin{aligned} L(\beta \mid \underline{x}) &= f_{\underline{X}}(\underline{x}) \overset{X_i \, indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \overset{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) \\ &= \prod_{i=1}^{n} \left[3\beta x_i^2 (1-x_i^3)^{\beta-1} \right] \\ &= 3^n \beta^n \prod_{i=1}^{n} \left[x_i^2 (1-x_i^3)^{\beta-1} \right], \quad \beta > 0 \end{aligned}$$

  Passo 2 — Função de log-verosimilhança

$$\ln L(\beta \mid \underline{x}) = n\ln(3) + n\ln(\beta) + 2\sum_{i=1}^{n} \log(x_i) + (\beta-1)\sum_{i=1}^{n} \ln\left(1-x_i^3\right), \quad \beta > 0$$

  Passo 3 — Maximização

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

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

$$\begin{cases} \frac{n}{\hat{\beta}} + \sum_{i=1}^{n} \ln\left(1-x_i^3\right) = 0 \\ -\frac{n}{\hat{\beta}^2} < 0 \end{cases} \Leftrightarrow \begin{cases} \hat{\beta} = -\frac{n}{\sum_{i=1}^{n} \ln\left(1-x_i^3\right)} \\ -\frac{n}{\hat{\beta}^2} < 0 \quad \text{(prop. verdadeira)} \end{cases}$$

  Passo 4 — Estimador de MV de β

$$EMV(\beta) = -\frac{n}{\sum_{i=1}^{n} \ln\left(1-X_i^3\right)}$$

Pergunta 7	2 valores

A gestora de uma unidade fabril decidiu estimar a probabilidade p de um parafuso de alta resistência escolhido ao acaso da produção diária não satisfazer as especificações de segurança. Para o efeito, selecionou uma amostra casual de 100 desses parafusos, dos quais 4 não satisfizeram as especificações de segurança.

Determine um intervalo de confiança aproximado a 95% para a probabilidade p.

A: [0.0394, 0.0406] **B**: [0.0016, 0.0784] **C**: [0.0078, 0.0722] **D**: [0.0392, 0.0408]

- **V.a. de interesse**

  $X = \begin{cases} 1, & \text{parafuso de alta resistência não satisfaz especificações de segurança} \\ 0, & \text{caso contrário} \end{cases}$

- **Situação**

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

  p DESCONHECIDO

- **Obtenção de IC aproximado para** p

  Passo 1 - Seleção da v.a. fulcral para p

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

  Passo 2 — Obtenção dos quantis de probabilidade

  Como $(1-\alpha)\times 100\% = 95\% \Leftrightarrow \alpha = 0.05$, lidaremos com os quantis seguintes:

  $$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(0.975) \overset{tabela/calc.}{=} -1.9600 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) = 1.9600. \end{cases}$$

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

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

  $$P\left[a_\alpha \le (\bar{X}-p)/\sqrt{\bar{X}(1-\bar{X})/n} \le b_\alpha\right] \simeq 1-\alpha$$

  $$P\left[\bar{X} - b_\alpha \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \le p \le \bar{X} - a_\alpha \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \simeq 1-\alpha$$

  Passo 4 — Concretização

  A expressão geral do intervalo aproximado de confiança a $(1-\alpha)\times 100\%$ para p é

  $$IC_{(1-\alpha)\times 100\%}(p) \simeq \left[\bar{x} \pm \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right].$$

  Ao termos em conta que $n = 100$, $\alpha = 0.05$, $\Phi^{-1}(1-\alpha/2) = 1.9600$ e $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{4}{100} = 0.04$,

  $$\begin{aligned} IC_{95\%}(p) &\simeq \left[0.04 - 1.9600 \times \sqrt{\frac{0.04\times(1-0.04)}{100}}, 0.04 + 1.9600 \times \sqrt{\frac{0.04\times(1-0.04)}{100}}\right] \\ &\simeq [0.0016, 0.0784]. \end{aligned}$$

Pergunta 8	2 valores

Um engenheiro físico avalia a distância percorrida no lançamento de um projétil sob novas condições de atrito. Efetuados 15 registos casuais da distância percorrida, obtiveram-se uma média e uma variância (corrigida) amostrais iguais a 9.8 e 1.2 (respetivamente).

Considerando tal distância uma variável aleatória com distribuição normal com valor esperado μ e desvio padrão σ desconhecidos, confronte as hipóteses $H_0 : \mu = 10.2$ e $H_1 : \mu < 10.2$. Obtenha o valor-p (ou um intervalo para o valor-p) do teste. Para que níveis de significância não se deve rejeitar H_0?

- **V.a. de interesse**

  X = distância percorrida no lançamento de projétil sob novas condiçes de atrito

- **Situação**

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

  μ desconhecido

  σ DESCONHECIDO

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 10.2 \quad \text{vs.} \quad H_1 : \mu < \mu_0$

- **Estatística de teste**

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{a}{\sim}_{H_0} t_{n-1}$$

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

  O teste é unilateral inferior ($H_1 : \mu < \mu_0$), logo a região de rejeição de H_0 é do tipo $W = (-\infty, -c)$.

- **Valor-p e comentário**

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

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{9.8 - 10.2}{\sqrt{1.2/15}} \simeq -1.41 \\ valor-p &= P(T < t \mid H_0) = F_{t_{(n-1)}}(t) \simeq F_{t_{(14)}}(-1.41) = 1 - F_{t_{(14)}}(1.41) \overset{calc.}{\simeq} 0.0902, \end{aligned}$$

  é suposto:

  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq valor-p \simeq 9.02\%$, designadamente aos níveis usuais de significâncias de 1% e 5%)[;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > valor-p \simeq 9.02\%$, nomeadamente ao n.u.s. de 10%].

  [Alternativamente, poderíamos recorrer às tabelas de quantis da distribuição t-Student e adiantar um intervalo para o valor-p:

$$\begin{aligned} F^{-1}_{t_{(14)}}(0.9) = 1.345 &< |t| = 1.41 < 1.523 = F^{-1}_{t_{(14)}}(0.925) \\ 0.9 &< F_{t_{(14)}}(1.41) < 0.925 \\ 1 - 0.925 &< 1 - F_{t_{(14)}}(0.52) < 1 - 0.9 \\ 0.075 &< valor-p < 0.1. \end{aligned}$$

  Assim, é suposto:

  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 7.5\%$, designadamente aos níveis usuais de significâncias de 1% e 5%)[;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 10\%$, nomeadamente ao n.u.s. de 10%.]

Pergunta 9	2 valores

A distribuição de Lomax, com função densidade de probabilidade dada por $f_0(x) = \frac{3}{(x+1)^4}$, para $x > 0$, foi apontada como adequada para descrever o tamanho de ficheiros para certa categoria de dados digitais (hipótese H_0).

Com base no teste de ajustamento do qui-quadrado e na tabela de frequências seguinte referente a 100 destes ficheiros selecionados casualmente, teste a adequabilidade de tal distribuição, ao nível de significância de 2.5%.

Tamanho de ficheiro (MB)	$(0,0.25]$	$(0.25,0.5]$	$(0.5,0.75]$	$(0.75,1]$	$(1,+\infty)$
Frequência absoluta observada	52	26	4	6	12
Frequência absoluta esperada sob H_0	48.80	21.57	10.97	6.16	12.5

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

  X = tamanho do ficheiro de dados

  $$f_0(x) = \begin{cases} \frac{3}{(x+1)^4}, & x > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

- **Hipóteses**

  $H_0 : f_X(x) \equiv f_0(x) \quad \text{vs.} \quad H_0 : f_X(x) \not\equiv f_0(x)$

- **N.s.**

  $\alpha_0 = 2.5\%$

- **Estatística de teste**

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

  onde: k = número 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

  $$c = F^{-1}_{\chi^2_{(k-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(5-1)}}(1-0.025) = F^{-1}_{\chi^2_{(4)}}(0.975) \stackrel{tabela/calc.}{=} 11.14.$$

- **Decisão**

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

i	Classe	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	$(0,0.25]$	52	48.8	$\frac{(52-48.8)^2}{48.8} \simeq 0.210$
2	$(0.25,0.5]$	26	21.57	0.910
3	$(0.5,0.75]$	4	10.97	4.429
4	$(0.75,1]$	6	6.16	0.004
5	$(1,+\infty)$	12	12.5	0.020
		$\sum_{i=1}^k o_i = n = 100$	$\sum_{i=1}^k E_i = n = 100$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i} \simeq 5.573$

  Como $t = 5.573 \not\in W = (11.14, +\infty)$, não devemos rejeitar H_0 ao n.s. $\alpha_0 = 2.5\%$.

Pergunta 10	2 valores

Dados relativos à percentagem de determinado tipo de madeira (x) em polpa de papel e a respetiva resistência à tração (Y) do papel resultante conduziram aos seguintes valores:

$$\textstyle\sum_{i=1}^{18} x_i = 78, \quad \sum_{i=1}^{18} x_i^2 = 511, \quad \sum_{i=1}^{18} y_i = 706, \quad \sum_{i=1}^{18} y_i^2 = 41\,209, \quad \sum_{i=1}^{18} x_i y_i = 4\,467,$$

onde $[\min_{i=1,\ldots,18} x_i, \max_{i=1,\ldots,18} x_i] = [1, 10]$.

Com base no modelo de regressão linear simples, $Y = \beta_0 + \beta_1 x + \epsilon$, e nos dados fornecidos, estime $E(Y \mid x = 7)$. Obtenha o coeficiente de determinação do modelo estimado e interprete o valor obtido.

- [**Modelo de RLS e hipóteses de trabalho**

  Y = resistência à tração (v.a. resposta)

  x = percentagem de determinado tipo de madeira (variável explicativa)

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

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

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

- **Estimativas de MQ dos parâmetros desconhecidos β_0 e β_1**

  Importa notar que

  - $n = 18$
  - $\sum_{i=1}^n x_i = 78$

    $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{78}{18} = 4.(3)$

    $\sum_{i=1}^n x_i^2 = 511$

    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 511 - 18 \times 4.(3)^2 = 173.0$
  - $\sum_{i=1}^n y_i = 706$

    $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{706}{18} = 39.(2)$

    $\sum_{i=1}^n y_i^2 = 41\,209$

    $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 41\,209 - 18 \times 39.(2)^2 = 13518.(1)$
  - $\sum_{i=1}^n x_i y_i = 4\,467$

    $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 4\,467 - 18 \times 4.(3) \times 39.(2) = 1\,407.(6)$.

  Logo,

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{1\,407.(6)}{173} \simeq 8.136802 \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \simeq 39.(2) - 8.136802 \times 4.(3) \simeq 3.962751.\end{aligned}$$

- **Estimativa pedida**

$$\hat{E}(Y \mid x = 7) \quad = \quad \hat{\beta}_0 + \hat{\beta}_1 \times 7 \simeq 3.962751 + 8.136802 \times 7 \simeq 60.920362.$$

- **Coeficiente de determinação**

$$r^2 = \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} = \frac{(1\,407.(6))^2}{173.0 \times 13518.(1)} \simeq 0.847301$$

- **Interpretação coeficiente de determinação**

  Cerca de 84.73% 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 bem ao conjunto de dados.